



SQ SEQUENCE 91 AA; 10075 MW; 44825 CN; 17.3%; Score 77; DB 2; Length 91; Best Local Similarity 38.5%; Pred. No. 1.77e+01; Matches 10; Conservative 9; Mismatches 6; Indels 1; Gaps 1;

Db 19 APASAPSYSDTTPCCPAYIAR-PLP 43  
QY 20 SSASPRKFLNTVTCCTATVTRIPLS 45

RESULT 2  
ID US-08-467-123B-5 STANDARD; PRT; 91 AA.  
XX  
AC xxxxxx  
XX  
DT  
XX  
DE Sequence 5, Application US/08467123B  
XX Sequence 5, Application US/08467123B  
CC Sequence 5, Application US/08467123B  
CC Patent No. 5945506  
CC GENERAL INFORMATION:  
CC APPLICANT: Coleman, Roger  
CC APPLICANT: Selhamer, Jeffrey J.  
CC TITLE OF INVENTION: CHEMOKINE EXPRESSED IN FETAL SPLEEN,  
CC ITS PRODUCTION AND USES  
CC NUMBER OF SEQUENCES: 9  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Incyte Pharmaceuticals, Inc.  
CC STREET: 3174 Porter Drive  
CC CITY: Palo Alto  
CC STATE: CA  
CC COUNTRY: USA  
CC ZIP: 94304  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Diskette  
CC COMPUTER: IBM Compatible  
CC OPERATING SYSTEM: DOS  
CC SOFTWARE: FastSeq Version 1.5  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/467,123B  
CC FILING DATE: 06-JUN-1995  
CC CLASSIFICATION: 530  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/375,346  
CC FILING DATE: 19-JAN-1995  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Billings, Lucy J.  
CC REGISTRATION NUMBER: 36,749  
CC REFERENCE/DOCKET NUMBER: PF-0026-1 DIV  
CC TELEPHONE: 415-555-0555  
CC TELEFAX: 415-845-4166  
CC INFORMATION FOR SEQ ID NO: 5:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 91 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: peptide  
CC NAME/KEY: Protein  
CC LOCATION: 1..68  
SQ SEQUENCE 91 AA; 10075 MW; 44825 CN;

Query Match 17.3%; Score 77; DB 2; Length 91;  
Best Local Similarity 38.5%; Pred. No. 1.77e+01;  
Matches 10; Conservative 9; Mismatches 6; Indels 1; Gaps 1;

Db 19 APASAPSYSDTTPCCPAYIAR-PLP 43  
QY 20 SSASPRKFLNTVTCCTATVTRIPLS 45

RESULT 3  
ID US-08-660-542-21 STANDARD; PRT; 91 AA.  
XX  
AC xxxxxx  
XX  
DT  
XX  
DE Sequence 21, Application US/08660542  
XX Sequence 21, Application US/08660542  
CC Patent No. 5932703  
CC GENERAL INFORMATION:  
CC APPLICANT: Godiska, Ronald  
CC APPLICANT: Gray, Patrick W.  
CC TITLE OF INVENTION: MACROPHAGE DERIVED CHEMOKINE AND CHEMOKINE  
CC ANALOGS  
CC NUMBER OF SEQUENCES: 32  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
CC STREET: 6300 Sears Tower, 233 South Wacker Drive  
CC CITY: Chicago  
CC STATE: Illinois  
CC COUNTRY: United States of America  
CC ZIP: 60606-6402  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/660,542  
CC FILING DATE:  
CC CLASSIFICATION: 514  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/558,658  
CC FILING DATE: 16-NOV-1995  
CC APPLICATION NUMBER: 08/479,620  
CC FILING DATE: 07-JUN-1995  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Gass, David A.  
CC REGISTRATION NUMBER: 38,153  
CC REFERENCE/DOCKET NUMBER: 27866/33318  
CC TELEPHONE: 312/474-6300  
CC TELEFAX: 312/474-0448  
CC TELEX: 25-3856  
CC INFORMATION FOR SEQ ID NO: 21:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 91 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: peptide  
CC NAME/KEY: misc feature  
CC OTHER INFORMATION: "RANTES"  
CC FEATURE:  
CC NAME/KEY: Protein  
CC LOCATION: 1..68  
SQ SEQUENCE 91 AA; 9990 MW; 44818 CN;

Query Match 17.3%; Score 77; DB 2; Length 91;  
Best Local Similarity 38.5%; Pred. No. 1.77e+01;  
Matches 10; Conservative 9; Mismatches 6; Indels 1; Gaps 1;

Db 19 APASAPSYSDTTPCCPAYIAR-PLP 43  
QY 20 SSASPRKFLNTVTCCTATVTRIPLS 45

RESULT 4  
ID US-08-480-449-21 STANDARD; PRT; 91 AA.





CC ADDRESSEE: Incyte Pharmaceuticals, Inc.

CC STREET: 3174 Porter Drive

CC CITY: Palo Alto

CC STATE: CA

CC COUNTRY: U.S.

CC ZIP: 94304

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Diskette

CC COMPUTER: IBM Compatible

CC OPERATING SYSTEM: DOS

CC SOFTWARE: FastSeq Version 1.5

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/798,143

CC FILING DATE: 10-FEB-1997

CC CLASSIFICATION: 536

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: 08/347,492

CC FILING DATE: 29-NOV-1994

CC APPLICATION NUMBER: 08/303,241

CC FILING DATE: 07-SEP-1994

CC APPLICATION NUMBER: 08/320,011

CC FILING DATE: 05-OCT-1994

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Luther, Barbara J

CC REGISTRATION NUMBER: 33,954

CC REFERENCE/DOCKET NUMBER: PF-0024

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: 415-855-0555

CC TELEFAX: 415-852-0195

CC INFORMATION FOR SEQ ID NO: 12:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 91 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: peptide

CC IMMEDIATE SOURCE:

CC LIBRARY: GENBANK

CC CLONE: GI 134510

CC SEQUENCE 91 AA; 10075 MW; 44825 CN;

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC COUNTRY: USA

CC ZIP: 94304

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Diskette

CC COMPUTER: IBM Compatible

CC OPERATING SYSTEM: DOS

CC SOFTWARE: FastSeq Version 1.5

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/375,346A

CC FILING DATE: 19-JAN-1995

CC CLASSIFICATION: 435

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER:

CC FILING DATE:

CC ATTORNEY/AGENT INFORMATION:

CC NAME: LUTHER, BARBARA J.

CC REGISTRATION NUMBER: 33,954

CC REFERENCE/DOCKET NUMBER: PF-0026 US

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (415) 855-0555

CC TELEFAX: (415) 855-0572

CC TELEX:

CC INFORMATION FOR SEQ ID NO: 5:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 91 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: peptide

CC HYPOTHETICAL: NO

CC ANTI-SENSE: NO

CC FRAGMENT TYPE: internal

CC ORIGINAL SOURCE:

CC SEQUENCE 91 AA; 10075 MW; 44825 CN;

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

RESULT	9	
ID	PCT-US92-09487-4	STANDARD; PRT; 263 AA.

Sequence 4, Application PC/TUS9209487

APPLICATION NUMBER: US 01/181,567

```

Query Match      17.3%;      Score 77;  DB 3;  Length 263
Best Local Similarity 26.3%;  Pred. No. 1.77e+01;
Matches 10;  Conservative 13;  Mismatches 14;  Indels

```

Db	24	TLPSHKVYDIPLLYSTISDSRRFILLDLTSAYETIS	61
Qv	4	TLVFSRRCIEPLTLTKSSASPRF-LTNVTGCCFATVT	40

RESULT	10	
ID	US-08-646-360-4	STANDARD; PRT; 263 AA.

AC  
XXXXXXXX

Sequence 4: Application US/03646360

Sequence 4, Application US/08646360

Patent No. 5837491  
GENERAL INFORMATION

APPLICANT: Better, Marc D

APPLICANT: Carroll, Stephen E.  
APPLICANT: Studnika, Gary M.

TITLE OF INVENTION:	IMMUNOTOX
TITLE OF INVENTION:	PROTEIN

TITLE OF INVENTION: FLOTTING  
 NUMBER OF SEQUENCES: 173

CORRESPONDENCE ADDRESS:  
ADDRESSEE. McAndrews Held

STREET: 500 West Madison St

CITY: Chicago  
STATE: Illinois

COUNTRY: USA

ZIP: 60661  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS

SOFTWARE: PatentIn Release  
CURRENT APPLICATION DATA:

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/6

FILING DATE: 13-MAY-1996  
CLASSIFICATION: 530

CLASSIFICATION: C50  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US99  
FILING DATE: 12-MAY-1999

·PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/012,512  
FILING DATE: 12-MAY-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/79  
FILING DATE: 09-DEC-1992

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PG 07/0

APPLICATION NUMBER: US 01/79  
FILING DATE: 19-JUN-1992

PRIOR APPLICATION DATA:



CC	INFORMATION FOR SEQ ID NO: 4
CC	SEQUENCE CHARACTERISTICS:
CC	LENGTH: 263 amino acids
CC	TYPE: amino acid
CC	TOPOLOGY: linear
CC	MOLECULE TYPE: protein
SQ	SEQUENCE 263 AA; 29601MW; 366

Query Match 17.3%; Score 77; DB 1; Length 263;  
Best Local Similarity 26.3%; Pred. No. 1.77e+01;  
Matches 10; Conservative 13; Mismatches 14; Indels

```
Db      24  TLPFSHKVYDIPLLYSTISDSRRFILLDLTSYAYETIS 61
        ||| : || : : | : | : | : | : | :
Qv      4  TLVQSRECFIFPLTFKSSASPRIF-LTNVTGCCFAVT 40
```

RESULT 13  
ID US-07-901-707-4 STANDARD; PRT; 263 AA.  
XX  
AC xxxxxx

Sequence 4, Application US/07901707  
Sequence 4, Application US/07901707

GENERAL INFORMATION:  
 APPLICANT: Bernhardt, Susan L.  
 APPLICANT: Better, Marc D.  
 APPLICANT: Carroll, Steve F.  
 APPLICANT: Lane, Julie A.  
 TITLE OF INVENTION: Materials Comprising and Methods of  
 Composition and Use for Ribosome-In  
 NUMBER OF SEQUENCES: 57

CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &  
CC ADDRESSEE: Bicknell  
CC STREET: Two First National Plaza, 20 South Clark  
CC STREET: Street  
CC CITY: Chicago  
CC STATE: Illinois  
CC COUNTRY: USA  
CC ZIP: 60603  
CC

CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/07/901.707  
CC FILING DATE: 19920619  
CC

CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/787,567  
CC FILING DATE: 04-NOV-1991  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: No. 5376546and, Greta E.  
CC REGISTRATION NUMBER: 35,302  
CC REFERENCE/DOCKET NUMBER: 27129/30910  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (312) 346-5750  
CC TELEFAX: (312) 984-5750  
CC TELEX: 25-3856  
CC

CC INDEX: 23-3830  
 CC INFORMATION FOR SEQ ID NO: 4:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 263 amino acids  
 CC TYPE: AMINO ACID  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: protein  
 CC SEQUENCE 263 AA; 29601 MW; 362575 CM;

	every Match	17.3%	Score 77;	DB 1;	Length 263;
	Best Local Similarity	26.3%	Pred. No. 1.77e+01;		
	Matches	10;	Mismatches	14;	Indels
Db	24	TLPSHKKYVDIPLLYSTISDSRRFFILLDLTSYAYETIS	51		
Qv	4	TLVCSRFICPLPLAFKKSASPKF-LNVTGCCFAT	40		

RESULT 14  
ID US-08-477-484B-4 STANDARD; PRT; 263 AA.

AC  
XX  
XX  
D<sup>r</sup>

Sequence 4, Application US/08477484B  
Sequence 4, Application US/08477484A  
Sequence 4, Application US/08477484A

CC  
CC  
CC  
CC  
CC  
CC

FACILE NO. 3750899

GENERAL INFORMATION:

APPLICANT: Better, Marc D.

APPLICANT: Carroll, Stephen F.

CC	APPLICANT: Studnika, Gary M.
CC	TITLE OF INVENTION: Immunotoxins
CC	TITLE OF INVENTION: Proteins

CC NUMBER OF SEQUENCES: 169  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: McAndrews, Held & M

STREET: 500 West Madison Street  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA

CC COUNTRY: USA  
CC ZIP: 60661  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk

```

CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.

```

CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/477,  
CC FILING DATE: 07-JUN-1995  
CC CLASSIFICATION: 520

CC CLASSIFICATION: 330  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/425,  
CC FILING DATE: 18-APR-1995

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/064,  
FILING DATE: 12-MAY-1993

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/988,  
FILING DATE: 09-DEC-1992

CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/901,  
CC FILING DATE: 19-JUN-1992  
CC PRIOR APPLICATION DATA:  
CC

APPLICATION NUMBER: US 07/787,  
FILING DATE: 04-NOV-1991  
ATTORNEY/AGENT INFORMATION:

CC NAME: McNicholas, Janet M.  
CC REGISTRATION NUMBER: 32,918  
CC REFERENCE/DOCKET NUMBER: 110222

TELEPHONE: 312/707-8889  
TELEFAX: 312/707-9155  
TELEF. 650 388-1248

CC  
CC  
CC  
CC  
CC

TELE: 050 300 1240  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 263 amino acids

CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein



```

Query Match      20.0%; Score 89; DB 2; Length 111;
Best Local Similarity 35.3%; Pred.No. 3.12e-02;
Matches         18; Conservative 17; Mismatches 11; Indels 5; Gaps 4;

Db    12 ALVHNHRENT-PPTLNFSA-QRIYFS--SGCTLSHLFLRLHLHS-KMLTAAEK 57
QY    6 TLVFECRCFICPLIFTKSSASPRKELNVTGGCFATVTRIPLSNKVLTAVDNR 54
      :|::||| |::|| ::|| :: || |::|| ::
      :|::||| |::|| ::|| ::|| ::|| ::|| ::
      :|::||| |::|| ::|| ::|| ::|| ::|| ::

RESULT 2
ENTRY   A38144             #type complete
TITLE   adenosine receptor Al - Bovine
ORGANISM #formal_name Bos primigenius taurus #common_name cattle
DATE    04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change
        29-Jan-1999
ACCESSIONS A38144; S20390; S22213
REFERENCE  A38144
AUTHORS   Olah, M.E.; Ren, H.; Ostrowski, J.; Jacobson, K.A.; Stiles,
           G.L.
#journal  J. Biol. Chem. (1992) 267:10764-10770
#title    Cloning, expression, and characterization of the unique
           bovine Al adenosine receptor. Studies on the ligand binding
           site by site-directed mutagenesis.
#crosso-references MWID:92268130
#accession A38144
#molecule_type mRNA

```

```

##residues      1-326 #label OLA
##cross-references GB:M86261; NID:g162597; PID:g162598
##experimental_source brain
##note          sequence extracted from NCBI backbone (NCBIN:103815,
                  NCBI:P103816)
REFERENCE
#authors        S20390
#journal        Tucker, A.L.; Linden, J.; Robeva, A.S.; D'Angelo, D.D.;
#title          Lynch, K.R.
#cross-references MUID:92201360
#accession      S20390
#status         nucleic acid sequence not shown
#molecule_type mRNA
#residues       1-100, 'M', 102-326 #label TUC
##cross-references EMBL:X63592
##experimental_source brain
#CLASSIFICATION #superfamily adenosine receptor A1
#FEATURES       adenylylate cyclase inhibitor; G protein-coupled receptor;
                  glycoprotein; transmembrane protein
FEATURE         148,159 #binding site carbohydrate (Asn) (covalent) #status
                  predicted
SUMMARY         #length 326 #molecular-weight 36579 #checksum 6464
Query Match    19.3%; Score 86; DB 2; Length 326;
Best Local Similarity 31.8%; Pred. No. 9.47e-02;
Matches 14; Conservative 13; Mismatches 14; Indels 3; Gaps 3;
Db 111 IPLRYKTVTPRAVVAITCGWILSFVVGITPMFGWNNLSAVER 154
QY 14 FPLTFKSSASPRKFLNTVGTGC-CFA-TVTRIPLSN-KVLTAVD 54

RESULT 3
ENTRY
TITLE         S56958 #type complete
ALTERNATE_NAMES probable membrane protein YJL175w - yeast (Saccharomyces
                  cerevisiae)
ORGANISM       hypothetical protein J0502
#formal_name Saccharomyces cerevisiae
DATE           08-Jul-1995 #sequence_revision 08-Sep-1995 #text_change
14-Nov-1997
ACCESSIONS     S56958
REFERENCE
#authors       Obermaier, B.; Piravandi, E.; Rinke, M.; Domdey, H.
#submission    submitted to the Protein Sequence Database, September 1995
#accession     S56958
#molecule_type DNA
#status        1-170 #label OBE
#residues      ##cross-references EMBL:Z49451; NID:g1008372; PID:g1008374; MIPS:YJL175w
#map_position  10L
#keywords       transmembrane protein
SUMMARY        #length 170 #molecular-weight 17731 #checksum 7245
Query Match    19.1%; Score 85; DB 2; Length 170;
Best Local Similarity 39.3%; Pred. No. 1.36e-01;
Matches 11; Conservative 7; Mismatches 9; Indels 1; Gaps 1;
Db 81 IVGCCFATVVIASLSRFGTVASSFR 108
QY 31 VTGCCFATVTRIPLS-NKVLTAVDRLR 57

RESULT 4
ENTRY
TITLE         VCFVER #type complete
CONTAINS      env polyprotein - Rous sarcoma virus (strain Prague C)
#formal_name gp37; coat protein gp85
ORGANISM       #formal_name Rous sarcoma virus
DATE           01-Sep-1981 #sequence_revision 17-Dec-1982 #text_change
04-Oct-1996
ACCESSIONS     A03996

```

```

REFERENCE
#authors        A00632
#submission     Schwartz, D.; Tizard, R.; Gilbert, W.
#note          submitted to the Nucleic Acid Sequence Database, September
                  1982
#accession      A03996
#molecule_type genomic RNA
#residues       1-603 #label SCH
#note          as a result of base variations, a different version of
                  this sequence may exist having 17-Glu, 134-Ser,
                  158-Ser, 334-Thr, 383-Thr, 392-Val, 522-Glu, 541-Leu,
                  and 567-Val
GENETICS
#gene          env
#CLASSIFICATION #superfamily type C retrovirus env polyprotein
#KEYWORDS      polyprotein
#FEATURE        65-603 #product coat protein gp85 #status predicted #label P85\
                  406-603 #product coat protein gp37 #status predicted #label P37
SUMMARY        #length 603 #molecular-weight 65660 #checksum 5554
Query Match    18.9%; Score 84; DB 1; Length 603;
Best Local Similarity 35.5%; Pred. No. 1.96e-01;
Matches 11; Conservative 10; Mismatches 9; Indels 1; Gaps 1;
Db 304 PFSTANTGSGN-LGNVSGCGCEPITILPLG 333
QY 15 PLTFKSSASPRKFLNTVGTGCCFATVTRIPLS 45

RESULT 5
ENTRY
TITLE         T00094 #type complete
ALTERNATE_NAMES endostyle-specific protein - sea squirt (Cliona intestinalis)
ORGANISM       #formal_name Cliona intestinalis
DATE           22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change
22-Jan-1999
ACCESSIONS     T00094
REFERENCE
#authors       Ogasawara, M.; Satoh, N.
#submission    submitted to the EMBL Data Library, February 1998
#description   isolation and characterization of endostyle-specific genes in
                  the Ascidian Cliona intestinalis.
#accession     T00094
#status        preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues      1-650 #label OGA
#cross-references EMBL:AB010895; NID:d1179651; PID:d1025751
#experimental_source endostyle
SUMMARY        #length 650 #molecular-weight 75483 #checksum 3537
Query Match    18.9%; Score 84; DB 2; Length 650;
Best Local Similarity 31.6%; Pred. No. 1.96e-01;
Matches 12; Conservative 10; Mismatches 12; Indels 4; Gaps 4;
Db 344 YKTRYCFVRYCLOPKFCFNPKY-TGLKTCVFPVAVR 380
QY 7, FQSRFC-I-FPL-TFKSSASPRKFLNTVGTGCCFATVTR 41

RESULT 6
ENTRY
TITLE         S70364 #type complete
ALTERNATE_NAMES cannabinoid receptor CB2, peripheral - mouse
ORGANISM       G protein-coupled receptor CB2
#formal_name Mus musculus #common_name house mouse
DATE           21-Apr-1997 #sequence_revision 09-May-1997 #text_change
17-Mar-1999
ACCESSIONS     S70364; S54163
REFERENCE
#authors       Shire, D.; Calandra, B.; Rinaldi-Carmona, M.; Oustric, D.;
                  Pessegue, B.; Bonnin-Cabanne, O.; le Fur, G.; Caput, D.;
                  Ferrara, P.
#journal        Biochim. Biophys. Acta (1996) 1307:132-136
#title          Molecular cloning, expression and function of the murine CB2
                  peripheral cannabinoid receptor.

```



ORGANISM #variety

•

```

#cross-references EMBL:Z50754; NID:G944824; PID:G944825
CLASSIFICATION #superfamily Shigella toxin chain A
SUMMARY #length 319 #molecular-weight 35610 #checksum 6388

Query Match 17.5%; Score 78; DB 2; Length 319;
Best Local Similarity 34.4%; Pred. No. 1.63e+00;
Matches 11; Conservative 8; Mismatches 11; Indels 2; Gaps 2;

Db 1 MKCIL-FKWLCLL-LGFSVSYSTREFTIDLS 30
QY 1 MKCTLVQSFRCIFPLTFKSSASPRKFLTNVT 32

RESULT 14
ENTRY #type fragment
TITLE mitochondrial solute carrier protein homolog - rat (fragment)
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 03-Aug-1992 #sequence_revision 03-Aug-1992 #text_change
09-Apr-1998
ACCESSIONS B40141
REFERENCE #authors Zarrilli, R.; Oates, E.L.; McBride, O.W.; Lerman, M.I.; Chan, J.Y.; Santisteban, P.; Ursini, M.V.; Notkins, A.L.; Kohn, L.D.
#journal Mol. Endocrinol. (1989) 3:1498-1508
#title Sequence and chromosomal assignment of a novel cDNA identified by immunoscreening of a thyroid expression library: similarity to a family of mitochondrial solute carrier proteins.
#cross-references MUID:90114217.
#accession B40141
#molecule_type mRNA
#residues 1-326 #label ZAR
#cross-references GB:M32973
#note the authors translated the codon CGA for residue 254 as Gly

CLASSIFICATION #superfamily ADP,ATP carrier protein repeat homology
KEYWORDS duplication; mitochondrion; transmembrane protein
FEATURE
33-121 #domain ADP,ATP carrier protein repeat homology #label ACP1\
130-220 #domain ADP,ATP carrier protein repeat homology #label ACP2\
240-326 #domain ADP,ATP carrier protein repeat homology #label (fragment) #label ACP3
SUMMARY #length 326 #checksum 6952

Query Match 17.5%; Score 78; DB 2; Length 326;
Best Local Similarity 48.1%; Pred. No. 1.63e+00;
Matches 13; Conservative 4; Mismatches 7; Indels 3; Gaps 3;

Db 38 RSFLAGTAGCC-AKTTVAPLDRVKVL 63
QY 25 RKFLT-NVTGCCFATVTRPLSN-KVL 49

RESULT 15
ENTRY #type complete
TITLE mitochondrial solute carrier protein homolog - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 03-Aug-1992 #sequence_revision 03-Aug-1992 #text_change
09-Apr-1998
ACCESSIONS A40141
REFERENCE #authors Zarrilli, R.; Oates, E.L.; McBride, O.W.; Lerman, M.I.; Chan, J.Y.; Santisteban, P.; Ursini, M.V.; Notkins, A.L.; Kohn, L.D.
#journal Mol. Endocrinol. (1989) 3:1498-1508
#title Sequence and chromosomal assignment of a novel cDNA identified by immunoscreening of a thyroid expression library: similarity to a family of mitochondrial solute carrier proteins.
#cross-references MUID:90114217

```

```

#accession A40141
#molecule_type mRNA
#residues 1-349 #label ZAR
#cross-references GB:M31659; NID:G188697; PID:G386960
#note the authors translated the codon CGA for residue 251 as Gly

COMMENT The intron information here reflects apparent unspliced or aberrantly spliced introns in the mRNA sequence. It is not based on genomic DNA.

GENETICS
#map_position 10
#introns 44/1; 141/3
CLASSIFICATION #superfamily ADP,ATP carrier protein repeat homology
KEYWORDS duplication; mitochondrion; transmembrane protein
FEATURE
33-121 #domain ADP,ATP carrier protein repeat homology #label ACP1\
127-217 #domain ADP,ATP carrier protein repeat homology #label ACP2\
237-329 #domain ADP,ATP carrier protein repeat homology #label ACP3
SUMMARY #length 349 #molecular-weight 38328 #checksum 7111

Query Match 17.5%; Score 78; DB 2; Length 349;
Best Local Similarity 48.1%; Pred. No. 1.63e+00;
Matches 13; Conservative 4; Mismatches 7; Indels 3; Gaps 3;

Db 38 RSFLAGTAGCC-AKTTVAPLDRVKVL 63
QY 25 RKFLT-NVTGCCFATVTRPLSN-KVL 49

Search completed: Wed Dec 22 10:19:54 1999
Job time : 10 secs.

```

\*\*\*\*\*  
MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Wed Dec 22 10:20:11 1999; MasPar time 4.57 Seconds  
Tabular output not generated. 364.887 Million cell updates/sec  
\*\*\*\*\*  
Title: >US-09-209-961-7  
Description: (1-59) from US09209961.pep  
Perfect Score: 445  
Sequence: 1 MKCTLVQSFRCIFPLTFKS.....TRIPLSNKVLTAVDRLRCP 59  
Scoring table: PAM 150  
Gap 11  
Searched: 77977 seqs, 28268293 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: swiss-prot37  
1:swissprot  
Statistics: Mean 36.740; Variance 52.488; scale 0.700  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.  
SUMMARIES

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Wed Dec 22 10:20:11 1999; MasPar time 4.57 Seconds  
Tabular output not generated. 364.887 Million cell updates/sec  
\*\*\*\*\*

Title: >US-09-209-961-7  
Description: (1-59) from US09209961.pep  
Perfect Score: 445  
Sequence: 1 MKCTLVQSFRCIFPLTFKS.....TRIPLSNKVLTAVDRLRCP 59  
Scoring table: PAM 150  
Gap 11  
Searched: 77977 seqs, 28268293 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: swiss-prot37  
1:swissprot  
Statistics: Mean 36.740; Variance 52.488; scale 0.700  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.  
SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description	Pred. No.
1	89	20.0	111	1	Y2PU_ECOLI	VERY HYPOTHETICAL 13.2	4.07e-03
2	86	19.3	326	1	AAIR_BOVIN	ADENOSINE A1 RECEPTOR	1.46e-02
3	85	19.1	170	1	YJRS_YEAST	HYPOTHETICAL 17.7 KD P	2.21e-02
4	84	18.9	603	1	ENV_RSVF	ENV POLYPROTEIN [CONTA	3.35e-02
5	83	18.7	347	1	CB2R_MOUSE	CANNABINOID RECEPTOR 2	5.06e-02
6	79	17.8	3421	1	TEGU_HSVB	LARGE TEGUMENT PROTEIN	2.54e-01
7	78	17.5	322	1	GDC_RAT	GRAVE'S DISEASE CARRIE	3.77e-01
8	78	17.5	332	1	GDC_HUMAN	GRAVE'S DISEASE CARRIE	3.77e-01
9	78	17.5	403	1	ZXDB_HUMAN	ZINC FINGER X-LINKED P	3.77e-01
10	78	17.5	457	1	ZXDB_HUMAN	ZINC FINGER X-LINKED P	3.77e-01
11	77	17.3	91	1	SISD_HUMAN	T-CELL SPECIFIC RANTES	5.57e-01
12	77	17.3	286	1	RIP2_MOMBA	RIPOSOOME-INACTIVATING	5.57e-01
13	77	17.3	330	1	GDC_BOVIN	GRAVE'S DISEASE CARRIE	5.57e-01
14	77	17.3	306	1	NUSW_RHUN	NADH-UBIQUINONE OXIDOR	5.57e-01
15	76	17.1	326	1	AAIR_HUMAN	ADENOSINE A1 RECEPTOR	8.20e-01
16	76	17.1	400	1	HOFC_ECOLI	PROTEIN TRANSPORT PROT	8.20e-01
17	76	17.1	684	1	TGLA_HUMAN	PROTEIN-GLUTAMINE GLUT	8.20e-01
18	75	16.9	319	1	SLTA_BP933	SHIGA-LIKE TOXIN II SU	1.20e+00
19	75	16.9	912	1	KPCM_HUMAN	PROTEIN KINASE C, MU T	1.20e+00
20	75	16.9	1887	1	YDHA_SCHPO	HYPOTHETICAL 192.5 KD	1.20e+00
21	74	16.6	308	1	Y117_HELPY	HYPOTHETICAL PROTEIN H	1.76e+00
22	74	16.6	531	1	UL95_HCMVA	PROTEIN UL95	1.76e+00
23	74	16.6	639	1	YAJ9_SCHPO	HYPOTHETICAL 74.4 KD P	1.76e+00

ID	Y2PU_ECOLI	STANDARD;	PRT;	111 AA.
AC	P24254;			
DT	01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)			
DT	01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)			
DE	VERY HYPOTHETICAL 13.2 KD PROTEIN IN PUTA-PUTP INTERGENIC REGION.			
OS	ESCHERICHIA COLI.			
OC	BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;			
OC	ESCHERICHIA			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-K12;			
RX	MEDLINE; 88142554.			
RA	NAKAO T., YAMATO I., ANRAKU Y.;			
RT	"Nucleotide sequence of putC, the regulatory region for the put			
RT	regulon of Escherichia coli K12."			
RL	MOL. GEN. GENET. 210:364-368(1987).			
CC	!- CAUTION: THIS SEQUENCE, ACCORDING TO THE EMBL DATABASE (K. RUDD)			
CC	IS PROBABLY NOT A REAL PROTEIN; THEREFORE THIS ENTRY WILL PROBABLY			
CC	BE DELETED IN FUTURE RELEASES.			
CC	-----			
CC	This Swiss-Prot entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; M35174; G147431.			
DR	EMBL; X05553; G42601;			
DR	PIR; S06385; S06385.			
RW	HYPOTHETICAL PROTEIN.			
SQ	SEQUENCE 111 AA; 13203 MW; 6E3E921D CRC32;			

ALIGNMENTS

Query Match	20.08;	Score 89;	DB 1;	Length 111;
Best Local Similarity	35.3%;	Pred. No. 4.07e-03;		
Matches	18;	Conservative 17;	Mismatches 11;	Indels 5;
				Gaps 4;
DB	12	ALVTHNRFNT-PFTLNESA-ORHYFS--SGCTLSHFLRLHLS-KMLTAAEK 57		
Qy	4	TLVFSRFCIFPLTFKSASPRFLTNVTCCEATVTRIPLSNKLTAVD 54		
RESULT	2			
ID	AAIR_BOVIN	STANDARD;	PRT;	326 AA.



CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))

DR EMBL; V01197; G61703;  
 DR PIR; A03996; VCFVER;  
 DR HSSP; P03385; IMOF;  
 KW COAT PROTEIN; POLYPROTEIN; TRANSMEMBRANE; SIGNAL;  
 FT SIGNAL 1 64  
 FT CHAIN 65 603 SURFACE PROTEIN GP85.  
 FT CHAIN 406 603 MEMBRANE PROTEIN GP37.  
 FT TRANSMEM 557 577 POTENTIAL.  
 SQ SEQUENCE 603 AA; 65660 MW; C80D2F1A CRC32;

Query Match 18.9%; Score 84; DB 1; Length 603;  
 Best Local Similarity 35.5%; Pred. No. 3.35e-02;  
 Matches 11; Conservative 10; Mismatches 9; Indels 1; Gaps 1;

Db 304 PFSFTANTGSGN-LGNVSGCGCEPITILPLG 333  
 QY 15 PLTFKSSASPRKFLTNVTCGFATVTRIPLS 45

RESULT 5  
 ID CB2R\_MOUSE STANDARD; PRT; 347 AA.  
 AC P47936;  
 DT 01-FEB-1996 (REL. 33, CREATED)  
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
 DE CANNABINOID RECEPTOR 2 (CB2) (CB-2).  
 GN CNR2.

OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN-C57BL/6; TISSUE-SPLEEN;  
 RX MEDLINE; 96283804.  
 RA SHIRE D., CALANDRA B., RINALDI-CARMONA M., OUSTRIC D., PESSEQUE B.,  
 RA CABANNE O., LE FUR G., CAPUT D., FERRARA P.;  
 RT "Molecular cloning, expression and function of the murine CB2  
 RT peripheral cannabinoid receptor."  
 RL BIOCHIM. BIOPHYS. ACTA 1307:132-136(1996).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC STRAIN-129/SV;  
 RA BUCKLEY N.E., BONNER T.I.;  
 RL SUBMITTED (FEB-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [3]

RP SEQUENCE FROM N.A.  
 RC STRAIN-NFS/N;  
 RX MEDLINE; 97404692.  
 RA VALK P.J.M., HOL S., VANKAN Y., IHLE J.N., ASKEW D., JENKINS N.A.,  
 RA GILBERT D.J., COPELAND N.G., DE BOTH N.J., LOWENBERG B., DELWEL R.;  
 RT "The genes encoding the peripheral cannabinoid receptor and alpha-l-  
 RT fucosidase are located near a newly identified common virus  
 RT integration site, Evill1."  
 RL J. VIROL. 71:6796-6804(1997).  
 CC -!- FUNCTION: INVOLVED IN CANNABINOID-INDUCED CNS EFFECTS. COULD BE A  
 CC RECEPTOR FOR ANANDAMIDE.

CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))

DR EMBL; X86405; G791082;  
 DR EMBL; U21681; G710428;

DR EMBL; X93168; E213356;  
 DR GCRDB; GCR\_1704;  
 DR GCRDB; GCR\_1725;  
 DR GCRDB; GCR\_2451;  
 DR MGD; MGI:104650; CNR2.  
 DR PROSITE; PS00237; G\_PROTEIN\_RECEPTOR; 1.  
 DR PFAM; PF00001; 7tm\_1; 1.

KW G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN.  
 FT DOMAIN 1 33  
 FT TRANSMEM 34 59  
 FT DOMAIN 60 71  
 FT TRANSMEM 72 92  
 FT DOMAIN 93 104  
 FT TRANSMEM 105 129  
 FT DOMAIN 130 149  
 FT TRANSMEM 150 172  
 FT DOMAIN 173 188  
 FT TRANSMEM 189 214  
 FT DOMAIN 215 246  
 FT TRANSMEM 247 267  
 FT DOMAIN 268 279  
 FT TRANSMEM 280 301  
 FT DOMAIN 302 347  
 FT CARBOHYD 11 11  
 FT CONFLICT 54 64  
 FT CONFLICT 126 126  
 FT CONFLICT 272 272  
 FT CONFLICT 332 332  
 SQ SEQUENCE 347 AA; 38213 MW; 753AD117 CRC32;

Query Match 18.7%; Score 83; DB 1; Length 347;  
 Best Local Similarity 34.5%; Pred. No. 5.06e-02;  
 Matches 19; Conservative 12; Mismatches 20; Indels 4; Gaps 4;

Db 84 SVIFACNFVIFHV-FHGVDNSAIFILKI-GSVTMTFT-ASVGSLLITAVDRYL-C 134  
 QY 4 TLVFCRSCFIPLTFKSSASPRKFLTNVTCGFATVTRIPLSNKVLTAVDRLSRC 58

RESULT 6

ID TEGU\_HVHB STANDARD; PRT; 3421 AA.  
 AC F28955;  
 DT 01-DEC-1992 (REL. 24, CREATED)  
 DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)  
 DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)  
 DE LARGE TEGUMENT PROTEIN.  
 GN 24.  
 OS EQUINE HERPESVIRUS TYPE 1 (STRAIN AB4P) (EHV-1).  
 OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;  
 OC ALPHAHERPESVIRINAE; VARICELLOVIRUS.

RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 92295566.  
 RA TELFORD E.A.R., WATSON M.S., MCBRIDE K., DAVISON A.J.;  
 RT "The DNA sequence of equine herpesvirus-1."  
 RL VIROLOGY 189:304-316(1992).

CC -!- FUNCTION: TEGUMENT PROTEIN.

CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,  
 CC EHV-1 24, EBV BFLF1, HVS-1 64, VZV 22, AND HCMV UL48.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))

DR EMBL; M86564; G330816;  
 DR PIR; G36797; WZBEB6.  
 DR HSSP; P04002; IMFA.  
 SQ SEQUENCE 3421 AA; 367078 MW; 9F15166B CRC32;

Query Match 17.8%; Score 79; DB 1; Length 3421;  
Best Local Similarity 30.8%; Pred. No. 2:54e-01;  
Matches 12; Conservative 14; Mismatches 11; Indels 2; Gaps 2;

DB 3373 FRSTLSAMALL-IAACR-TIVRLRATRVLDINRSL 3409  
QY 18 FKSSAPRFLTNVTCGCFATVTRIPLSN-KVLAVDRSL 56

RESULT 7  
ID GDC\_RAT STANDARD; PRT; 322 AA.  
AC P16261;  
DT 01-AUG-1990 (REL. 15, CREATED)  
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)  
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)  
DE GRAVE'S DISEASE CARRIER PROTEIN (GDC) (MITOCHONDRIAL SOLUTE CARRIER  
DE RATINUS NORVEGICUS (FRAGMENT).  
OS RATTUS NORVEGICUS (RAT).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.  
RN [1]

SEQUENCE FROM N.A.  
RC TISSUE-THYROID;  
RX MEDLINE; 90114217.  
RA ZARRILLI R., OATES E.L., MCBRIDE O.W., LERMAN M.I., CHAN J.Y.,  
RA SANTISTEBAN P., URSINI M.V., NOTKINS A.L., KOHN L.D.;  
RT "Sequence and chromosomal assignment of a novel cDNA identified by  
RT immunoscreening of a thyroid expression library: similarity to a  
RT family of mitochondrial solute carrier proteins.";  
RL MOL. ENDOCRINOL. 3:1498-1508(1989).  
CC -!- FUNCTION: NOT KNOWN.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL  
CC INNER MEMBRANE (POTENTIAL).  
CC -!- DOMAIN: COMPOSED OF THREE HOMOLOGOUS DOMAINS.  
CC -!- SIMILARITY: BELONGS TO THE MITOCHONDRIAL CARRIER FAMILY.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; M32973; G205529;  
DR PROSITE; PS00215; MITOCH\_CARRIER; 3.  
DR PFAM; PF00153; mito\_carr; 1  
KW MITOCHONDRION; TRANSMEMBRANE; TRANSPORT; REPEAT.  
FT REPEAT 1 123  
FT REPEAT 124 222  
FT REPEAT 223 322  
FT NON\_TER 322 322  
SQ SEQUENCE 322 AA; 35056 MW; 346AFDA6 CRC32;

Query Match 17.5%; Score 78; DB 1; Length 322;  
Best Local Similarity 48.1%; Pred. No. 3.77e-01;  
Matches 13; Conservative 4; Mismatches 7; Indels 3; Gaps 3;

DB 38 RSFLAGSIAGCC-AKTIVAPLDRVKVL 63  
QY 25 RKFLT-NVTGCCFATVTRIPLSN-KVL 49

RESULT 8  
ID GDC\_HUMAN STANDARD; PRT; 332 AA.  
AC P16260;  
DT 01-AUG-1990 (REL. 15, CREATED)  
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)  
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)  
DE GRAVE'S DISEASE CARRIER PROTEIN (GDC) (MITOCHONDRIAL SOLUTE CARRIER  
DE PROTEIN HOMOLOG).  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;

OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-THYROID;  
RX MEDLINE; 90114217.  
RA ZARRILLI R., OATES E.L., MCBRIDE O.W., LERMAN M.I., CHAN J.Y.,  
RA SANTISTEBAN P., URSINI M.V., NOTKINS A.L., KOHN L.D.;  
RT "Sequence and chromosomal assignment of a novel cDNA identified by  
RT immunoscreening of a thyroid expression library: similarity to a  
RT family of mitochondrial solute carrier proteins.";  
RL MOL. ENDOCRINOL. 3:1498-1508(1989).  
CC -!- FUNCTION: NOT KNOWN.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL  
CC INNER MEMBRANE (POTENTIAL).  
CC -!- DOMAIN: COMPOSED OF THREE HOMOLOGOUS DOMAINS.  
CC -!- DISEASE: POSSIBLE ROLE IN GRAVES' DISEASE.  
CC -!- SIMILARITY: BELONGS TO THE MITOCHONDRIAL CARRIER FAMILY.  
CC -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 320  
CC ONWARD DUE TO A FRAMESHIFT.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; M31659; G386960; ALT\_FRAME.  
DR PIR; A40141; A40141.  
DR MIM; L39080;  
DR PROSITE; PS00215; MITOCH\_CARRIER; 3.  
DR PFAM; PF00153; mito\_carr; 1.  
KW MITOCHONDRION; TRANSMEMBRANE; TRANSPORT; REPEAT.  
FT REPEAT 1 123  
FT REPEAT 124 219  
FT REPEAT 220 332  
SQ SEQUENCE 332 AA; 36235 MW; 7E1DB8DC CRC32;

Query Match 17.5%; Score 78; DB 1; Length 332;  
Best Local Similarity 48.1%; Pred. No. 3.77e-01;  
Matches 13; Conservative 4; Mismatches 7; Indels 3; Gaps 3;

DB 38 RSFLAGSIAGCC-AKTIVAPLDRVKVL 63  
QY 25 RKFLT-NVTGCCFATVTRIPLSN-KVL 49

RESULT 9  
ID ZXDB\_HUMAN STANDARD; PRT; 403 AA.  
AC P98169;  
DT 01-OCT-1996 (REL. 34, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE ZINC FINGER X-LINKED PROTEIN ZXDB (FRAGMENT).  
GN ZXDB.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
RN [1]

RP SEQUENCE FROM N.A.  
RC TISSUE-BRAIN;  
RX MEDLINE; 94093547.  
RA GREIG G.M., SHARP C.B., CARREL L., WILLARD H.F.;  
RT "Duplicated zinc finger protein genes on the proximal short arm of

the human X chromosome: isolation, characterization and  
 RT X-inactivation studies.;  
 RL HUM. MOL. GENET. 2:1611-1618(1993).  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
 CC -!- TISSUE SPECIFICITY: UBIQUITOUS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC

EMBL; L14788; G292933;  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2; 8.  
 DR PFAM; PF00096; zf-C2H2; 9.  
 DR HSP; P08047; ISP2.  
 KW ZINC-FINGER; METAL-BINDING; DNA-BINDING; REPEAT; NUCLEAR PROTEIN.

FT NON\_TER 1 1  
 FT DOMAIN <1 266 ZINC-FINGERS.  
 FT ZN\_FING <1 22 C2H2-TYPE.  
 FT ZN\_FING 28 52 C2H2-TYPE.  
 FT ZN\_FING 58 80 C2H2-TYPE.  
 FT ZN\_FING 87 111 C2H2-TYPE.  
 FT ZN\_FING 118 142 C2H2-TYPE.  
 FT ZN\_FING 148 172 C2H2-TYPE.  
 FT ZN\_FING 178 202 C2H2-TYPE.  
 FT ZN\_FING 208 232 C2H2-TYPE.  
 FT ZN\_FING 241 266 C2H2-TYPE.  
 FT NON\_TER 403 403  
 SQ SEQUENCE 403 AA; 44403 MW; 57AD6326 CRC32;

Query Match 17.5%; Score 78; DB 1; Length 403;  
 Best Local Similarity 25.4%; Pred. No. 3.77e-01;  
 Matches 15; Conservative 19; Mismatches 21; Indels 4; Gaps 3;

Db 125 CSKQYD-KACRLKHLRSHTGERPFLCDGCGWNTSMKSL-LRHKKHDDRRMCP 181  
 QY 3 CTLVFQSFRCIFPLTFKSSASPRKFLNTVGTCC--FATVTRPLSNKVLTAVDRLRCP 59

RESULT 10  
 ID ZXDA\_HUMAN STANDARD; PRT; 457 AA.  
 AC P98168;  
 DT 01-OCT-1996 (REL. 34, CREATED)  
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE ZINC FINGER X-LINKED PROTEIN ZXDA (FRAGMENT).  
 GN ZXDA.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 94093547.  
 RA GREIG G.M., SHARP C.B., CARREL L., WILLARD H.F.;  
 RT "Duplicated zinc finger protein genes on the proximal short arm of  
 RT the human X chromosome: isolation, characterization and  
 RT X-inactivation studies.";  
 RL HUM. MOL. GENET. 2:1611-1618(1993).  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
 CC -!- TISSUE SPECIFICITY: UBIQUITOUS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC

EMBL; L14787; G292931;  
 DR PROSITE; PS00026; ZINC\_FINGER\_C2H2; 8.  
 DR PFAM; PF00096; zf-C2H2; 9.  
 DR HSP; P08047; IS22.  
 KW ZINC-FINGER; METAL-BINDING; DNA-BINDING; REPEAT; NUCLEAR PROTEIN.  
 FT NON\_TER 1 1  
 FT DOMAIN <1 266 ZINC-FINGERS.  
 FT ZN\_FING <1 22 C2H2-TYPE.  
 FT ZN\_FING 28 52 C2H2-TYPE.  
 FT ZN\_FING 58 80 C2H2-TYPE.  
 FT ZN\_FING 87 111 C2H2-TYPE.  
 FT ZN\_FING 118 142 C2H2-TYPE.  
 FT ZN\_FING 148 172 C2H2-TYPE.  
 FT ZN\_FING 178 202 C2H2-TYPE.  
 FT ZN\_FING 208 232 C2H2-TYPE.  
 FT ZN\_FING 241 266 C2H2-TYPE.  
 FT NON\_TER 457 AA; 49978 MW; 10CC47AE CRC32;

Query Match 17.5%; Score 78; DB 1; Length 457;  
 Best Local Similarity 25.4%; Pred. No. 3.77e-01;  
 Matches 15; Conservative 19; Mismatches 21; Indels 4; Gaps 3;

Db 125 CSKQYD-KACRLKHLRSHTGERPFLCDGCGWNTSMKSL-LRHKKHDDRRMCP 181  
 QY 3 CTLVFQSFRCIFPLTFKSSASPRKFLNTVGTCC--FATVTRPLSNKVLTAVDRLRCP 59

RESULT 11  
 ID SISD\_HUMAN STANDARD; PRT; 91 AA.  
 AC P13501;  
 DT 01-JAN-1990 (REL. 13, CREATED)  
 DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE T-CELL SPECIFIC RANTES PROTEIN PRECURSOR (SIS-DELTA) (SMALL INDUCIBLE  
 DE CYTOKINE A5).  
 GN SCYA5.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 88285659.  
 RA SCHALL T.J., JONGSTRA J., DYER B.J., JORGENSEN J., CLAYBERGER C.,  
 RA DAVIS M.M., KRENSKY A.M.;  
 RT "A human T cell-specific molecule is a member of a new gene family.";  
 RL J. IMMUNOL. 141:1018-1025(1988).  
 RN [2]  
 RP STRUCTURE BY NMR.  
 RX MEDLINE; 95352612.  
 RA CHUNG C.-W., COOKE R.M., PROUDFOOT A.E.I., WELLS T.N.C.;  
 RT "The three-dimensional solution structure of RANTES.";  
 RL BIOCHEMISTRY 34:9307-9314(1995).  
 RN [3]  
 RP STRUCTURE BY NMR.  
 RX MEDLINE; 95244456.  
 RA SKELTON N.J., ASPIRAS F., OGEE J., SCHALL T.J.;  
 RT "Proton NMR assignments and solution conformation of RANTES, a  
 RT chemokine of the C-C type.";  
 RL BIOCHEMISTRY 34:5329-5342(1995).

CC -!- FUNCTION: CHEMOATTRACTANT FOR BLOOD MONOCYTES, MEMORY T HELPER  
 CC CELLS AND EOSINOPHILS. CAUSES THE RELEASE OF HISTAMINE FROM  
 CC BASOPHILS AND ACTIVATES EOSINOPHILS.  
 CC -!- TISSUE SPECIFICITY: T-CELL AND MACROPHAGE SPECIFIC.  
 CC -!- INDUCTION: BY MITOGENS.  
 CC -!- SIMILARITY: BELONGS TO THE INTERCRINE BETA FAMILY (SMALL CYTOKINE  
 CC C-C) (CHEMOKINE CC).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC



or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

```
CC EMBL; M21121; G339421;
CC PIR; A28815; A28815.
CC PDB; 1HRJ; 14-OCT-96.
CC PDB; 1RPN; 03-JUN-95.
CC PDB; 1RTO; 03-JUN-95.
CC MIM; 187011; -.
CC PROSITE; PS00472; SMALL_CYTOKINES_CC; 1.
CC PFAM; PF00048; i18; 1.
CC CYTOKINE; CHEMOTAXIS; T-CELL; SIGNAL; INFLAMMATORY RESPONSE;
KW 3D-STRUCTURE. 1 23
FT CHAIN 24 91
FT DISULFID 33 57
FT DISULFID 34 73
FT SEQUENCE 91 AA; 10075 MW; 6BDFE23E CRC32;
Query Match 17.3%; Score 77; DB 1; Length 91;
Best Local Similarity 38.5%; Pred. No. 5.57e-01;
Matches 10; Conservative 9; Mismatches 6; Indels 1; Gaps 1;
Db 19 APASAPYSSDTPCCFAYIAR-PLP 43
Qy 20 SSASPRKFLTNVTCGFATVTRIPLS 45
RESULT 12
ID RIP2_MOMBA STANDARD; PRT; 286 AA.
AC F29339;
DT 01-DEC-1992 (REL. 24, CREATED)
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE RIBOSOME-INACTIVATING PROTEIN MOMORDIN II PRECURSOR (RRNA
DE N-GLYCOSIDASE) (EC 3.2.2.22).
OS MOMORDICA BALSAMINA (BITTER GOURD) (BALSAM PEAR).
OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; TRACHEOPHYTA;
OC EUPHYLLIPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC CUCURBITALES; CUCURBITACEAE; MOMORDICA.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SEED;
RX MEDLINE; 93027170.
RA ORTIGAO M., BETTER M.;
RT "Momordin II, a ribosome inactivating protein from Momordica
RT balsamina, is homologous to other plant proteins.";
RL NUCLEIC ACIDS RES. 20:4662-4662(1992).
CC -!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF THE N-GLYCOSIDIC BOND AT ONE
CC SPECIFIC ADENOSINE ON THE 28S RRNA.
CC -!- SIMILARITY: TO OTHER BACTERIAL AND PLANTS RIBOSOME-INACTIVATING
CC PROTEINS. BELONGS TO TYPE 1 RIP.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC EMBL; X66035; G387.
CC PIR; S26596; S26596.
CC PROSITE; PS00215; MITOCH_CARRIER; 3.
CC PFAM; PF00153; mito_carr; 1.
CC MITOCHONDRION; TRANSMEMBRANE; TRANSPORT; REPEAT.
FT REPEAT 1 121
FT REPEAT 122 217
FT REPEAT 218 330
FT SEQUENCE 330 AA; 36085 MW; 428468D4 CRC32;
Query Match 17.3%; Score 77; DB 1; Length 330;
Best Local Similarity 48.1%; Pred. No. 5.57e-01;
Matches 13; Conservative 3; Mismatches 8; Indels 3; Gaps 3;
Db 36 RSFLAGGIAGCC-AKTVAPLDVRKVL 61
Qy 25 RKFLTN-VTGCFFATVTRIPLSN-KVL 49
RESULT 14
ID NUSM_RHUN STANDARD; PRT; 606 AA.
AC Q96069;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 5 (EC 1.6.5.3).
GN MIND5 OR NDS OR NADH5.
OS RHINOCEROS UNICORNIS (GREATER INDIAN RHINOCEROS).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
```



\*\*\*\*\*  
 M P S R C H  
 (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Dec 22 10:20:37 1999; MasPar time 10.31 Seconds  
 312.331 Million cell updates/sec

Tabular output not generated.

Title: >US-09-209-961-7  
 Description: (1-59) from US09209961.pep  
 Perfect Score: 445  
 Sequence: 1 MKCTLVFQSRFCIFPLTFKS.....TRIPLSNKVLTAVDRLRCP 59

Scoring table: PAM 150  
 Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: sptrembl9  
 1:sp.archaea 2:sp.bacteria 3:sp.fungi 4:sp.human  
 5:sp.invertebrate 6:sp.mammal 7:sp.mhc 8:sp.organelle  
 9:sp.phage 10:sp.plant 11:sp.rodent 12:sp.unclassified  
 13:sp.vertebrate 14:sp.virus

Statistics: Mean 35.655; Variance 56.771; scale 0.628

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	445	100.0	59	14	P6.6.	1.02e-84
2	418	93.9	59	14	STRAIN PMWS PCV, COMPL	6.24e-78
3	89	20.0	1206	11	SYNAPTOJANIN 2 (FRAGME	2.13e-02
4	89	20.0	1248	11	SYNAPTOJANIN II	2.13e-02
5	84	18.9	402	14	ENVELOPE POLYPROTEIN (	1.48e-01
6	84	18.9	454	5	T27F6.5 PROTEIN.	1.48e-01
7	84	18.9	601	14	ENV-PR95 POLYPROTEIN P	1.48e-01
8	84	18.9	650	5	ENDOSTYLE-SPECIFIC.	1.48e-01
9	83	18.7	141	14	GENOME, PARTIAL SPOUEN	2.16e-01
10	82	18.4	319	2	SHIGA-LIKE TOXIN II SU	3.15e-01
11	80	18.0	323	5	F28A12.2 PROTEIN.	6.62e-01
12	80	18.0	345	14	(RAV-0 SUBGROUP E) GP-	6.62e-01
13	79	17.8	286	10	MAP30 (FRAGMENT).	9.56e-01
14	79	17.8	311	6	UNCOUPLING PROTEIN 3.	9.56e-01
15	79	17.8	326	10	GAMMA-GLUTAMYL HYDROLA	9.56e-01
16	79	17.8	592	2	PEPTIDASE, PUTATIVE.	9.56e-01
17	78	17.5	295	5	SIMILAR TO MOUSE CREB-	1.38e+00
18	78	17.5	313	14	POLYPROTEIN (FRAGMENT)	1.38e+00
19	78	17.5	319	2	(EC 3.2.2.22) (RRNA N-	1.38e+00
20	77	17.3	91	4	RANTES PRECURSOR.	1.98e+00

21	77	17.3	125	6	077766	M3 MUSCARINIC ACETYLYCH	1.98e+00
22	77	17.3	260	14	Q65555	UL34.	1.98e+00
23	77	17.3	501	5	Q16923	F25G6.7 PROTEIN.	1.98e+00
24	77	17.3	4976	2	087314	FXBC.	2.83e+00
25	76	17.1	304	5	Q45882	W03B1.7 PROTEIN.	2.83e+00
26	76	17.1	320	2	Q46050	(EC 3.2.2.22) (RRNA N-	2.83e+00
27	76	17.1	440	2	Q25161	NIFS-LIKE PROTEIN.	2.83e+00
28	76	17.1	443	5	P91351	CODED FOR BY C. ELEGAN	2.83e+00
29	76	17.1	648	10	Q41717	ACETOACTATE SYNTHASE	2.83e+00
30	76	17.1	684	4	Q16707	PROSTATE-SPECIFIC TRAN	2.83e+00
31	76	17.1	845	5	P91184	SIMILAR TO C. ELEGANS	2.83e+00
32	75	16.9	271	14	Q85474	ENVELOPE PROTEIN (FRAG	4.03e+00
33	75	16.9	301	5	Q22299	T07C5.4 PROTEIN.	4.03e+00
34	75	16.9	319	2	Q47642	(EC 3.2.2.22) (RRNA N-	4.03e+00
35	75	16.9	319	2	Q47636	(EC 3.2.2.22) (RRNA N-	4.03e+00
36	75	16.9	335	14	Q85475	ENVELOPE PROTEIN (FRAG	4.03e+00
37	75	16.9	454	5	P90792	D2030.7 PROTEIN.	4.03e+00
38	75	16.9	577	14	Q07453	ENV POLYPROTEIN (CONTA	4.03e+00
39	75	16.9	595	14	Q03819	ENV POLYPROTEIN (CONTA	4.03e+00
40	75	16.9	607	14	Q83134	PROTEIN KINASE C, MU (	4.03e+00
41	75	16.9	918	11	Q62101	ANKYRIN HOMOLOG.	4.03e+00
42	75	16.9	936	10	Q23292	1-EVIDENCE-PREDICTED B	4.03e+00
43	75	16.9	990	5	Q46086	SYNAPTOJANIN 2 ISOFORM	4.03e+00
44	75	16.9	1145	11	Q88399	POLYPROTEIN.	4.03e+00
45	75	16.9	2842	14	Q36178		

## ALIGNMENTS

RESULT 1  
 ID O92287 PRELIMINARY; PRT; 59 AA.  
 AC O92287;  
 DT 01-NOV-1998 (TREMBLREL. 08, CREATED)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE P6.6.  
 OS PORCINE CIRCOVIRUS.  
 OC VIRUSES; SSDNA VIRUSES; CIRCOVIRIDAE; CIRCOVIRUS.  
 K1 [1]  
 KP SEQUENCE FROM N.A.  
 RC STRAIN=412;  
 RA WANG L., WILLSON P., CHOW B., GIBBONS E., BABIUK L.;  
 "Emergence of a new porcine circovirus."  
 RL SUBMITTED (AUG-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: AF085695; G3668369;  
 SQ SEQUENCE 59 AA; 6603 MW; 016C3C5F CRC32;

Query Match 100.0%; Score 445; DB 14; Length 59;  
 Best Local Similarity 100.0%; Pred. No. 1.02e-84;  
 Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CU 1 MKCTLVFQSRFCIFPLTFKSSASPRKFLTNVTCGCFATVTRIPLSNKVLTAVDRLRCP 59  
 CY 1 MKCTLVFQSRFCIFPLTFKSSASPRKFLTNVTCGCFATVTRIPLSNKVLTAVDRLRCP 59

RESULT 2  
 ID 056125 PRELIMINARY; PRT; 59 AA.  
 AC 056125;  
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE STRAIN PMWS PCV, COMPLETE GENOME.  
 OS PORCINE CIRCOVIRUS.  
 OC VIRUSES; SSDNA VIRUSES; CIRCOVIRIDAE; CIRCOVIRUS.  
 K1 [1]  
 KP SEQUENCE FROM N.A.  
 RC STRAIN=PMWS PCV;  
 RA HAMEL A.L., LIN L.L., NAYAK G.P.S.;  
 "Nucleotide sequence of porcine circovirus associated with  
 postweaning multisystemic wasting syndrome in pigs."  
 J. VIROL. 72:5262-5267(1998).

RN SEQUENCE FROM N.A.  
 RP STRAIN-PWS PCV;  
 RA HAMEL A.L., LIN L.L., NAYAR G.P.S.;  
 RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-PORCINE CIRCOVIRUS TYPE II;  
 RX MEDLINE; 98418498.  
 RA MEEHAN B.M., MCNEILLY F.M., TODD D., KENNEDY S., JEWELL V.,  
 RA ELLIS J.A., HASSARD L.E., CLARK E.G., HAINES D.M., ALLAN G.M.;  
 RT "Characterisation of novel circovirus DNAs associated with wasting  
 RT syndromes in pigs";  
 RL J. GEN. VIROL. 79:2171-2199(1998).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-PORCINE CIRCOVIRUS TYPE II;  
 RA MEEHAN B.M.;  
 RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL; AF027217; G2689649; -  
 DR EMBL; AF055394; G3598835; -  
 DR EMBL; AF055391; G3598800; -  
 DR EMBL; AF055392; G3598812; -  
 DR EMBL; AF055393; G3598824; -  
 SQ SEQUENCE 59 AA; 6516 MW; 11076BA8 CRC32;  
 Query Match 93.9%; Score 418; DB 14; Length 59;  
 Best Local Similarity 94.9%; Pred. No. 6.24e-78;  
 Matches 56; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 Db 1 MCTLVFQSRFCIFPLTFKSSASPRKFLNTVGTCCATVTRPLSNKVLTAVDRLRCP 59  
 QY 1 MKCTLVFQSRFCIFPLTFKSSASPRKFLNTVGTCCATVTRPLSNKVLTAVDRLRCP 59

RESULT 3  
 ID O35404 PRELIMINARY; PRT; 1206 AA.  
 AC O35404;  
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE SYNAPTOJANIN 2 (FRAGMENT).  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
 OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J;  
 RX MEDLINE; 98113197.  
 RA KHVOTCHEV M., SUDHOF T.C.;  
 RT "Developmentally regulated alternative splicing in a novel  
 RT synaptojanin";  
 RL J. BIOL. CHEM. 273:2306-2311(1998).  
 DR EMBL; AF026123; G3478621; -  
 FT NON-TER 1  
 SQ SEQUENCE 1206 AA; 133879 MW; C63591C9 CRC32;  
 Query Match 20.0%; Score 89; DB 11; Length 1206;  
 Best Local Similarity 46.9%; Pred. No. 2.13e-02;  
 Matches 15; Conservative 7; Mismatches 8; Indels 2; Gaps 1;  
 Db 21 CLGELSLKSGVPLSLVLTGTC--TSVGRIP 50  
 QY 12 CIPPLTFKSSASPRKFLNTVGTCCFATVTRIP 43

RESULT 4  
 ID O55207 PRELIMINARY; PRT; 1248 AA.  
 AC O55207;  
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE SYNAPTOJANIN II.

OS RATTUS NORVEGICUS (RAT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
 OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 98049546.  
 RA NEMOTO Y., ARRIBAS M., HAFNER C., DECAMILLI P.;  
 RT "Synaptojanin 2, a novel synaptojanin isoform with a distinct  
 RT targeting domain and expression pattern";  
 RL J. BIOL. CHEM. 272:30817-30821(1997).  
 DR EMBL; U90312; G2708493; -  
 SQ SEQUENCE 1248 AA; 138274 MW; 4E57DB9E CRC32;  
 Query Match 20.0%; Score 89; DB 11; Length 1248;  
 Best Local Similarity 46.9%; Pred. No. 2.13e-02;  
 Matches 15; Conservative 7; Mismatches 8; Indels 2; Gaps 1;  
 Db 63 CLGELSLKSGVPLSLVLTGTC--TSVGRIP 92  
 QY 12 CIPPLTFKSSASPRKFLNTVGTCCFATVTRIP 43

RESULT 5  
 ID Q85484 PRELIMINARY; PRT; 402 AA.  
 AC Q85484;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE ENVELOPE POLYPROTEIN (FRAGMENT).  
 GN ENV.  
 OS ROUS SARCOMA VIRUS.  
 OC VIRUSES; RETROVIRUSES; RETROVIRIDAE; AVIAN TYPE C RETROVIRUSES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SCHMIDT-RUBIN SUBGROUP D;  
 RX MEDLINE; 88063012.  
 RA BOVA C.A., OLSEN J.C., SWANSTROM R.;  
 RT "The avian retrovirus env gene family: molecular analysis of host  
 RT range and antigenic variants";  
 RL J. VIROL. 62:75-83(1988).  
 DR EMBL; M22730; G825438; -  
 KW POLYPROTEIN; ENVELOPE PROTEIN.  
 FT NON-TER 1  
 FT CHAIN 53 >402 GLYCOPROTEIN 85.  
 FT NON-TER 402 402  
 SQ SEQUENCE 402 AA; 43565 MW; 9279511E CRC32;  
 Query Match 18.9%; Score 84; DB 14; Length 402;  
 Best Local Similarity 35.3%; Pred. No. 1.48e-01;  
 Matches 12; Conservative 10; Mismatches 11; Indels 1; Gaps 1;  
 Db 301 PISFTVNTGSGN-LGNVSGCCGEATILPLGAW 333  
 QY 15 PLTFKSSASPRKFLNTVGTCCFATVTRIPLSNKV 48

RESULT 6  
 ID O45869 PRELIMINARY; PRT; 454 AA.  
 AC O45869;  
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)  
 DE T27F6.5 PROTEIN.  
 GN T27F6.5.  
 OS CAENORHABDITIS ELEGANS.  
 OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;  
 OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA DOBSON R.;  
 RL SUBMITTED (NOV-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [2]  
 RP SEQUENCE FROM N.A.

Query Match	18.9%	Score 84;	DB 5;	Length 454;
Best Local Similarity	35.4%;	Pred. No. 1.48e-01;		
Matches	17;	Conservative	10;	Mismatches 17;
				Indels 4;
				Gaps 4;

```
.
Db      268 FHLGTKYSEALGAKFGQKPLDMCCFGIGVTRLLPAAIDLLSYSDKALR 315
...    | | | | | : : : : : ||| : : : : : ||| : : : : :
Qv     -14 FPLTFKSSAS-PRKFLTN-VTGCFAT-VTRI-PLSNKKVLPAVDRLSLR 57
```

RESULT	7	
ID	Q64984	PRELIMINARY; PRT: 601 AA.
AC	Q64984;	
DT	01-NOV-1996 (TREMBREL. 01, CREATED)	
DT	01-NOV-1996 (TREMBREL. 01, LAST SEQUENCE UPDATE)	
DT	01-NOV-1998 (TREMBREL. 08, LAST ANNOTATION UPDATE)	
DE	ENV-PR95 POLI PROTEIN PRECURSOR.	
GN	ENV.	
OS	ROUS SARCOMA VIRUS.	
OC	VIRUSES; RETROID VIRUSES; RETROVIRIDAE; AVIAN TYPE C RETROVIRUSES.	
[1]	SEQUENCE FROM N.A.	
RN	STRAIN-PRAGUE;	
RC	MEDLINE; 83155662.	
RX	SCHWARTZ D.E., TIZARD R., GILBERT W.;	
RT	"Nucleotide sequence of Rous sarcoma virus.";	
RL	CELL 32:853-869(1983).	
RN	[2]	
RN	SEQUENCE FROM N.A.	
RP	STRAIN-PRAGUE;	
RC	MEDLINE; 85124605.	
RX	BROOME S., GILBERT W.;	
RA	"Rous sarcoma virus encodes a transcriptional activator.";	
RT	CELL 40:537-546(1985).	
RN	[3]	
RN	SEQUENCE FROM N.A.	
RP	PETROPOULOS C.J.;	
RA	(IN) COFFIN J.M. (EDS.); RETROVIRUSES:757-757;	
RL	COLD SPRING HARBOR LABORATORY PRESS, NY, USA (1997).	
RL	[4]	
RN	SEQUENCE FROM N.A.	
RP		

RA		CHAPERONE C.		SUBMITTED (NOV-1987)	TO EMBL/GENBANK/DDBJ DATA BANKS.
DR	EMBL;	J02342; G210174;	- -		
DR	EMBL;	AF033808; G2801462;	- -		
DR	POLYPROTEIN.	KW			
FT	CHAIN	1	175	POTENTIAL.	
FT	CHAIN	63	601	POTENTIAL.	
FT	CHAIN	404	601	POTENTIAL.	
SQ	SEQUENCE	601 AA;	65316 MW;	6A2AF544	CRC32;

Query Match	18.9%;	Score 84;	DB 14;	Length 601;
Best Local Similarity	35.5%;	Pred. No. 1.48e-01;		
Matches	11;	Conservative	10;	Mismatches 9; Indels 1; Gaps 1;

```
Db      302 PFSTANCTGSN-LGNVSCCGEPITILPLG 331
      |::: : |:::| :|:::| :|:::|
```

QY		15.	PLTFKSSASPRKFLTNVTCGCCFATVTRIPLS	45	
RESULT	8				
ID	O44236	PRELIMINARY;	PRT;	650 AA.	
AC	O44236;				
.DT	01-JUN-1998	(TREMBUREL. 06,	CREATED)		
DT	01-JUN-1998	(TREMBUREL. 06,	LAST SEQUENCE UPDATE)		
DT	01-JUN-1998	(TREMBUREL. 06,	LAST ANNOTATION UPDATE)		
DE	ENDOSTYLE-SPECIFIC.				
OE	CIONA INTESTINALIS.				
OC	EUKARYOTA; METAZOA;	CHORDATA; UROCHORDATA; ASCIDIACEA; PHLEBOBRANCHIA;			
CC	CIONILAE; CIONA.				
RN	{1}				
RF	SEQUENCE FROM N.A.				
PC	TISSUE-ENDOSTYLE;				
MA	OCA SAWARA M., SATOH N.;				
NL	SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.				
DR	EMBL: AB010895; D1025751.				
SO	SEQUENCE : 650 AA; 75483 MW; 077A54DA CRC32;				
		18.9%; Score 84; DB 5; Length 650;			
		Best Local Similarity 31.6%;			
		Matches 12; Conservative 10; Mismatches 12; Indels 4; Gaps 4;			
Dd	34;	IKTRYCFVRICLOQFKFCFNPTKY-TGLKTCVFPAVTR	380		
	:	:   :	:	:	:
	:	:   :	:	:	:
	:	:   :	:	:	:
QY	7.	FQSRFC-I-FPL-TFKSSASPRKFLTNVTCGCCFATVTR	41		

Db 34d YKTRYCFVRYCLOPKFCFNPTKY-TGLKTCVFPAVR 380  
 : : | : : k : | : | : | : | :  
 QY 7.FQSRFC-I-FPL-TFKSSASPRKFLTNVTGCCFATVTR 41

RESULT	0	PRELIMINARY;	PRT;	141 AA.
ID	Q84563	AC		
DT	01-NOV-1996	(TREMBLREL, 01, CREATED)		
DT	01-NOV-1996	(TREMBLREL, 01, LAST SEQUENCE UPDATE)		
DT	01-NOV-1998	(TREMBLREL, 08, LAST ANNOTATION UPDATE)		
FE	GENOME, PARTIAL SEQUENCE			
FN	A349L.			
OS	PARAMETRICUM BURSARIA CHLORELLA VIRUS 1 (PBCV-1).			
OC	VIRUSES; DSNA VIRUSES, NO RNA STAGE; PHYCODNAVIRIDAE; PHYCODNAVIRUS.			
RN	SEQUENCE FROM H.A.			
RP	SEQUENCE			
RX	MEDLINE; 95133167.			
RA	LU Z., LI Y., SHANG Y., KUTISH G.F., ROCK D.L., VAN ETTEN J.L.			
RT	"Analysis of 45 kb of DNA located at the left end of the chlorella virus PBCV-1 genome."			
RL	VIROLOGY 206:339-352(1995).			
DR	EMBL; U42580; GI181512;			
SQ	SEQUENCE 141 AA; 16438 MW; DCEBBA2E CRC32;			

```

Query Match      18.7%   Score 83;   DB 14;   Length 141;
Best Local Similarity 38.1%   Pred. No. 2.10e-01;
Matches      8;   Conservative      9;   Mismatches      3;   Indels      1;   Gaps      1

```

RESULT	10			
ID	Q03037	PRELIMINARY;	PRT;	319 AA.
AC	Q03037			
UT	01-NOV-1996	(TREMBUREL. 01. CREATED)		
DT	01-NOV-1996	(TREMBUREL. 01. LAST SEQUENCE UPDATE)		
DT	01-NOV-1998	(TREMBUREL. 08. LAST ANNOTATION UPDATE)		
DE	SHIGA-LIKE TOXIN II SUBUNIT A PRECURSOR (VEROTOXIN 2 SUBUNIT A)			
DE	(RNA N-CYCLOSIDASE) (EC 3.2.2.22).			
GN	ST-T1A.			
OS	ESCHERICHIA COLI.			
OC	BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;			
OC	ESCHERICHIA.			
RN	[1]			
FP	SEQUENCE FROM N.A.			

RC STRAIN-OK3:H21;  
RA PATON A.W., PATON J.C., HEUZENROEDER M.W., GOLDWATER P.N.,  
RL MANNING P.A.;  
RM MICROB. PATHOG. 13:225-236(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-OL11:H-;  
RX MEDLINE; 94018566.  
RA PATON A.W., PATON J.C., MANNING P.A.;  
RT "Polymerase chain reaction amplification, cloning and sequencing of  
RL variant Escherichia coli Shiga-like toxin type II operons";  
RM MICROB. PATHOG. 15:77-82(1993).  
CC -!- FUNCTION: THE SUBUNIT A IS RESPONSIBLE FOR INHIBITING PROTEIN  
CC SYNTHESIS THROUGH THE CATALYTIC INACTIVATION OF 60S RIBOSOMAL  
CC SUBUNITS.  
CC -!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF THE N-GLYCOSIDIC BOND AT  
CC ONE SPECIFIC ADENOSINE ON THE 28S RRNA.  
CC -!- SUBUNIT: THE SHIGA-LIKE TOXINS CONTAIN A SINGLE A SUBUNIT AND  
CC MULTIPLE COPIES OF A B SUBUNIT.  
CC -!- SIMILARITY: TO ALPHA CHAINS OF SHIGA TOXIN, SHIGA-LIKE TOXIN I,  
CC TO THE A SUBUNIT OF THE PLANT TOXIN RICIN, AND TO OTHER PLANT  
CC RIBOSOME-INACTIVATING PROTEINS.  
DR EMBL; X65949; G49090;  
DR EMBL; L11078; G304951;  
DR PFAM; PF00161; RIP; 1.  
KW HYDROLASE; GLYCOSIDASE; TOXIN; TRANSMEMBRANE; SIGNAL.  
FT SIGNAL 1 22 POTENTIAL.  
FT CHAIN 23 319 SHIGA-LIKE TOXIN II SUBUNIT A.  
FT TRANSMEM 232 267 POTENTIAL.  
FT SITE 188 188 ESSENTIAL FOR ENZYMIC ACTIVITY.  
FT SITE 192 192 ESSENTIAL FOR ENZYMIC ACTIVITY.  
FT ACT\_SITE 189 189 BY SIMILARITY.  
FT DISULFID 263 282 POTENTIAL.  
FT VARIANT 198 198 R -> G (IN STRAIN OL11:H-).  
SQ SEQUENCE 319 AA; 35709 MW; F0331EA CRC32;  
Query Match 18.4%; Score 82; DB 2; Length 319;  
Best Local Similarity 36.7%; Pred. No. 3.15e-01;  
Matches 11; Conservative 7; Mismatches 10; Indels 2; Gaps 2;  
Db 1 MKCIL-FKWVLCIL-LGFSVSVSREPMID 28  
QY 1 MKCTLVQSRFCIFPLTFKSSASPRKFLTN 30  
RESULT 11  
ID Q22969 PRELIMINARY; PRT; 323 AA.  
AC Q22969;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE F28A12.2 PROTEIN.  
GN F28A12.2  
OS CAENORHABDITIS ELGANS.  
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEIA; RHABDITIA; RHABDITIDA;  
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE; 94150718.  
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,  
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAYELLO A., FULTON L.,  
RA GARDNER A., GREEN P., HANKINS T., HILLIER L., JIER M., JOHNSTON L.,  
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
RA LIGHTNING J., LLOYD C., McMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
RA PARSONS J., PERCY C., RIFEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
RA SHALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,  
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RL NATURE 368:32-38(1994).

AN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA SAMMONS L., MURRAY J.;  
RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBDJ DATA BANKS.  
DR EMBL; U64851; G2911709;  
SQ SEQUENCE 323 AA; 35770 MW; FCA220D4 CRC32;  
Query Match 18.0%; Score 80; DB 5; Length 323;  
Best Local Similarity 34.3%; Pred. No. 6.62e-01;  
Matches 12; Conservative 8; Mismatches 13; Indels 2; Gaps 2;  
Db 10 IFNRFNFSIPFHSIGIAPGRRTMNSIFCCFNS 44  
QY 6 VFQSRFCIF-PLTEKSSASP-RKELTNVTCGFAT 38  
RESULT 12  
ID Q85473 PRELIMINARY; PRT; 345 AA.  
AC Q85473;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE (RAV-0 SUBGROUP E) GP-85 (ENV) (FRAGMENT).  
OS ROUS SARCOMA VIRUS.  
OC VIRUSES; RETROID VIRUSES; RETROVIRIDAE; AVIAN TYPE C RETROVIRUSES.  
RN [1]  
RP SEQUENCE OF 270-345 FROM N.A.  
RA DORNER A.J.;  
RL SUBMITTED (JUL-1988) TO EMBL/GENBANK/DBDJ DATA BANKS.  
RN [2]  
RP SEQUENCE OF 1-269 FROM N.A.  
RX MEDLINE; 85083270.  
RA DORNER A.J., STOYE J.P., COFFIN J.M.;  
RT "Molecular basis of host range variation in avian retroviruses";  
RL J. VIROL. 53:32-39(1985).  
DR EMBL; M12172; G559652;  
FT NON\_TER 1 1  
SQ SEQUENCE 345 AA; 37411 MW; B26AE001 CRC32;  
Query Match 18.0%; Score 80; DB 14; Length 345;  
Best Local Similarity 37.9%; Pred. No. 6.62e-01;  
Matches 11; Conservative 10; Mismatches 7; Indels 1; Gaps 1;  
Db 244 PFSTANCTASN-LGNVSGCGCKTITLP 271  
QY 15-PLTFKSSASPRKFLTNVTCGFATVTRIP 43  
RESULT 13  
ID Q41257 PRELIMINARY; PRT; 286 AA.  
AC Q41257;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)  
DE MAP30 (FRAGMENT).  
GN MAP30.  
OS MOMORDICA CHARANTIA (BITTER GOURD) (BALSAM PEAR).  
OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRIOPHYTA; TRACHEOPHYTA;  
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;  
OC CUCURBITALES; CUCURBITACEAE; MOMORDICA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 95394347.  
RA LEE-HUANG S., HUANG P.L., CHEN H.C., HUANG P.L., BOURINBAIAI A.,  
RA HUANG H.I., KUNG H.F.;  
RT "Anti-HIV and anti-tumor activities of recombinant MAP30 from bitter  
RL melon";  
RL GENE 161:151-156(1995).  
DR EMBL; S79450; G1110597;  
DR PFAM; PF00161; RIP; 1.  
DR MENDEL; 15516; MOMCH; Rip; mn15516.  
FT NON\_TER 286

```

Query Match      17.8%  Score 799  DB 10; Length 326;
Best Local Similarity 39.4%  Pred. No. 9.56e-01;
Matches 13; Conservative 7; Mismatches 11; Indels

Db      191 C-LVQNQHRFGISQSFEGNIALSNFFKIVTTC 222
Qv      3  CTLVQSS-RFCIFPLTFTKSSASPRKELTNTVTC 34

```

Search completed: Wed Dec 22 10:20:51 1999  
Job time : 14 secs.

Query Match 17.8%; Score 79; DB 6; Length 311;  
Best Local Similarity 35.5%; Pred. No. 9.56e-01;  
Matches 11; Conservative 11; Mismatches 8; Indels

RESULT	15
ID	O65355
AC	PRELIMINARY; PRT: 326 RA.
AD	O65355;
DT DT	(TREMBLERL. 07, CREATED)
DT	01-AUG-1998 (TREMBLERL. 07, LAST SEQUENCE UPDATE)
DT DT	01-NOV-1998 (TREMBLERL. 08, LAST ANNOTATION UPDATE)
DE	GAMMA-GLUTAMYL HYDROLASE.
DS	ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC OC	EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; TRACHEOPHYTES; ROSALES; EUPHRYLOTES; SPERMATOPHYTES; MAGNOLIOPHYTES; EUDICOTYLEDONS; ROSELES; CAPRANALES; BRASSICACEAE; ARABIDOPSIS.
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-CV. COLOMBIA;
RX	MEDLINE; 97069645.
RA	HUANGPU J., PAK J.H., GRAHAM M.C., RICKLE S.A., GRAHAM J.S.;
RT RT	"Purification and molecular analysis of an extracellular gamma-glutamyl hydrolase present in young tissues of the soybean plant.";
RL	BIOCHEM. BIOPHYS RES. COMMUN 228:1-6(1996).

RA RICKLE S.A., LIU C.Y., XU H., GRAHAM J.S.;  
RL SUBMITTED (MAY-1998) TO EMBL/GENBANK/DBJ DATA BANKS.

\*\*\*\*\*  
WIPERH  
\*\*\*\*\*  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Wed Dec 22 10:22:13 1999; MasPar time 13.84 Seconds  
Tabular output not generated. 159.807 Million cell updates/sec

Title: >US-09-209-961-9  
Description: (1-104) from US09209961.pep  
Perfect Score: 807  
Sequence: 1 MVTIPPLVFRWFPVCGFRVC.....HSSRQVTPLSLRSRSTFNK 104  
Scoring table: PAM 150  
Gap 11

Searched: 170751 seqs, 21266608 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq35  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39

Statistics: Mean 29.035; Variance 113.207; scale 0.256

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

Result No.	Query			ID	Description	Pred. No.
	Score	Match	Length			
1	89	11.0	397	28	Human neurotactin.	9.65e+00
2	89	11.0	397	26	Novel human CX3C 397	9.65e+00
3	89	11.0	787	32	Homo sapiens Smoothen	9.65e+00
4	89	11.0	793	32	Rattus norvegicus Smo	9.65e+00
5	86	10.7	534	26	A. thaliana inorganic	1.69e+01
6	86	10.7	534	27	Arabidopsis thaliana	1.69e+01
7	84	10.4	440	2	Natural recombinant h	2.45e+01
8	84	10.4	440	25	Human lecithin-choles	2.45e+01
9	81	10.0	111	26	Novel human CX3C 111	4.23e+01
10	81	10.0	159	12	Tomv coat protein.	4.23e+01
11	81	10.0	372	28	N-methylamino acid ox	4.23e+01
12	81	10.0	956	26	HCWV UL105 open read	4.23e+01
13	81	10.0	956	26	Human cytomegalovirus	4.23e+01
14	80	9.9	348	12	POU domain transcript	5.07e+01
15	79	9.8	227	39	Fatty acid desaturase	6.08e+01
16	79	9.8	227	38	A putative delta-5 de	6.08e+01

SUMMARIES

17	79	9.8	227	37	W84146	Dictyostelium discoid	6.08e+01
18	75	9.8	345	31	W56756	Serotonin 5HT7 recept	6.08e+01
19	79	9.8	404	7	R34581	Mammalian serotonin r	6.08e+01
20	79	9.8	435	12	R58689	Rat REC20 serotonin r	6.08e+01
21	79	9.8	445	10	R54782	Human brain serotonin	6.08e+01
22	79	9.8	448	11	R57200	Rat 5HT6 receptor.	6.08e+01
23	79	9.8	450	19	R99427	Human lymphocyte spec	6.08e+01
24	79	9.8	451	29	W38426	Human multiple myelom	6.08e+01
25	79	9.7	439	34	W62858	Peniophora lycii 6-ph	7.27e+01
26	78	9.7	439	36	W74484	Peniophora phytase po	7.27e+01
27	77	9.5	109	34	W29636	Human secreted protei	8.69e+01
28	77	9.5	159	12	R60611	Tobamovirus coat prot	8.69e+01
29	77	9.5	456	28	W41695	Rabbit LPS-binding pr	8.69e+01
30	77	9.5	482	10	R53461	Lapine polysaccharide	8.69e+01
31	77	9.5	482	8	R41659	Lapine gram-negative	8.69e+01
32	77	9.5	482	27	W40813	Rabbit lipopolysaccha	8.69e+01
33	77	9.5	824	34	W68093	Human neuronal PAS do	8.69e+01
34	76	9.4	164	21	W11787	Early onset Alzheimer	1.04e+02
35	76	9.4	207	22	W15288	Mouse oligodendrocyte	1.04e+02
36	75	9.3	189	13	R66605	Flower style-specific	1.24e+02
37	75	9.3	189	21	W14913	Tomato S-ribonuclease	1.24e+02
38	75	9.3	221	34	W38578	S. pneumoniae stage V	1.24e+02
39	75	9.3	319	18	W00628	Protein kinase #1 con	1.24e+02
40	75	9.3	450	19	R99426	Murine lymphocyte spe	1.24e+02
41	75	9.3	534	30	W45748	Arabidopsis thaliana	1.24e+02
42	75	9.3	631	18	R77896	Bacterial transferrin	1.24e+02
43	75	9.3	631	30	W54128	H. influenzae strain	1.24e+02
44	75	9.3	631	28	W08970	Amino acid sequence o	1.24e+02
45	75	9.3	679	1	P95056	Swine parvovirus B ge	1.24e+02

ALIGNMENTS

PS-SHUT 1  
W34307 standard; Protein: 397 AA.  
W34307; 11-MAY-1998 (first entry)  
Human neurotactin.  
KW Neurotactin; human; chemokine; inflammatory disease; inflammation;  
KW neutrophil; chemotaxis; myelogenous leukaemia; polycythemia vera;  
KW hypermegakaryocytic disorder; chemoprotective;  
KW radioprotective; therapy; diagnosis; Bardet-Biedl syndrome.  
OS Homo sapiens.  
TS Key Location/Qualifiers  
FT Peptide 1..21  
FT Domain /label= Sig\_peptide  
FT Domain 22..341 /note= "extracellular domain"  
FT Domain 22..92 /note= "chemokine-like domain (Claim 5)"  
FT Domain 342..362 /note= "transmembrane domain"  
FT Domain 363..397 /note= "cytoplasmic domain"

W09742224-11-  
13-NOV-1997;  
06-MAY-1997; U07737  
05-MAY-1997; US-851160.  
07-MAY-1996; US-643798.  
(MILL-) MILLENNIUM PHARM INC.  
Pan Y:  
WPI: 97-556911/51.  
N-FSDB: T93210.  
DNA encoding murine and human neurotactin - useful for diagnosis and  
treatment of inflammatory disorders  
Claim 5; Page 76-78; 101pp; English.  
This protein is human neurotactin, a novel chemokine that stimulates  
chemotaxis of neutrophils. Its amino acid sequence was deduced from  
the isolated chemokine-like domain, but its overall structure  
distinguishes it from other chemokines. It is highly expressed in  
the normal human brain, and may play a role in Bardet-Biedl  
syndrome. Neurotactin polypeptides, including soluble neurotactin



RESULT	ID	Accession	Protein	397 AA
AC	W23345	standard	Protein	397 AA
DE	30-MAR-1998	(first entry)		
DE	Novel human CX3C	397 amino acid chemokine		
KW	Chemokine motif	CX3C chemokine; membrane bound CX3C chemokine;		
KW	proadherent; T cell; monocyte; soluble CX3C chemokine; arthritis;			
KW	proadhesive activity; inflammatory disease; compromised immune response;			
KW	neurogenic inflammation.			
KW	Homo sapiens.			
OS	Homo sapiens.			
FH	key			
FH	Peptide			
FT	1..24	/label= signal_sequence		
FT	/note= "putative"			
FT	25..111	/label= mature_protein		
FT	25..100	/note= "chemokine domain"		
FT	101..341	/note= "stalk region which possesses many potential glycosylation sites"		
FT	342..361	/note= "transmembrane region"		
FT	362..397	/note= "intracellular domain"		
FT	382	/note= "phosphorylation site"		
FT	392	/note= "Phosphorylation site"		
FT	WO9727299-A1.			
PN	31-JUL-1997.			
PD	23-JAN-1997.	U00293.		
PF	16-MAY-1996;	US-649006.		
PR	24-JAN-1996;	US-590828.		
PR	(SCHE ) SCHERING CORP.			
PA	Bazan JF, Schall TJ, Zlotnik A;			
PI	WPI: 97-393687/36.			
DR	N-PSDB: T64943.			
DR	New isolated mammalian CX3C chemokine genes - used to develop			
PT	products for the diagnosis and treatment of disorders involving CX3C			
PT	chemokine misregulation			
PS	Claim 18; Pages 75-76; 88pp; English.			

Query Match 11.0%; Score 89; DB 32; Length 787;  
 Best Local Similarity 27.9%; Pred. No. 9.65e+00;  
 Matches 12; Conservative 13; Mismatches 16; Indels 2; Gaps 2;

Db 457 flafgvlitfchfydfnqae-wersfrdyvlcqanvtigl 498  
 QY 39 YIGFPITLLHFAHFKFSQPAEIDFKRYV-LLCNHQNPAI 80

## RESULT 4

ID W37911 standard; Protein; 793 AA.  
 AC W37911;  
 DE Rattus norvegicus Smoothed protein.  
 KW Smoothed; treatment; cancer; vertebrae; vsmo; tissue;  
 KW regeneration; skin; lung; muscle; neural; bone; gut; disease;  
 KW ageing; trauma; detection; diagnosis; drug screening.  
 OS Rattus norvegicus.  
 FN W09814475-A1.  
 PD 09-APR-1998.  
 PF 29-SEP-1997; U17433.  
 PR 30-SEP-1996; US-720484.  
 PA (GETH ) GENENTECH INC.  
 PI De SAUVAGE FJ, Rosenthal A, Stone DM;  
 DR WPI: 98-240021/21.  
 DR N-PSDB; V29124.  
 PT New isolated vertebrate Smoothed proteins - is used to develop  
 PT products for treatment of cancers or for use in inducing  
 PT regeneration of tissue damaged by disease, ageing or trauma  
 PS Claim 6; Fig 1; 58pp; English.  
 CC The sequence is that of rat Smoothed protein (Smo).  
 CC Smo interacts with Hedgehog and Patched signalling molecules.  
 CC involved in cell proliferation and differentiation. The products  
 CC can be used to develop agents to provide protection from conditions  
 CC associated with constitutive activity of Smo or Hedgehog, including  
 CC some forms of cancer that may result from e.g. basal cell carcinoma,  
 CC basal cell nevus syndrome and pancreatic carcinoma. Smo agonists  
 CC can be used to induce the formation of, or enhance or stimulate  
 CC tissue regeneration, such as regeneration of skin tissue, lung  
 CC tissue, muscle (such as heart or skeletal muscle), neural tissue  
 CC (such as serotonergic neurons, motoneurons or striatal neurons),  
 CC bone tissue or gut tissue, which may be damaged by disease,  
 CC ageing or trauma. The products can also be used for detection,  
 CC diagnosis and drug screening.  
 SQ Sequence 793 AA;

Query Match 11.0%; Score 89; DB 32; Length 793;  
 Best Local Similarity 27.9%; Pred. No. 9.65e+00;  
 Matches 12; Conservative 13; Mismatches 16; Indels 2; Gaps 2;

Db 461 flafgvlitfchfydfnqae-wersfrdyvlcqanvtigl 502  
 QY 39 YIGFPITLLHFAHFKFSQPAEIDFKRYV-LLCNHQNPAI 80

## RESULT 5

ID W36137 standard; Protein; 534 AA.  
 AC W36137;  
 DE A. thaliana inorganic phosphate transporter protein IPT4.  
 KW Promoter; inorganic phosphate transporter gene; IPT4; expression;  
 KW regulation; transgenic plant; expression vector.  
 OS Arabidopsis thaliana.  
 FN J09252782-A.  
 PD 30-SEP-1997.  
 PF 25-MAR-1996; 094856.  
 PR 25-MAR-1996; JP-094856.  
 PA (MITS-) MITSUI GYOSAI SHOKUBUTSU BIO KENKYUSHO.  
 DR WPI: 97-530154/49.  
 DR N-PSDB; V01455.  
 PT A promoter for a phosphate transporter gene - useful for regulating  
 PT heterologous gene expression in transgenic plants  
 PS Example 2; Page 6-9; 12pp; Japanese.

CC This is the amino acid sequence of the Arabidopsis thaliana inorganic  
 CC phosphate transporter gene IPT4. The promoter region (bases 1-1875; see  
 CC also V01454) of the encoding gene can be used to regulate the level of  
 CC expression of a heterologous gene in a plant by introducing, into a  
 CC plant, a vector having the heterologous gene linked downstream region  
 CC from the promoter and then regulating extracellular phosphate levels.  
 SQ Sequence 534 AA;

Query Match 10.7%; Score 86; DB 26; Length 534;

Best Local Similarity 42.1%; Pred. No. 1.69e+01;  
 Matches 24; Conservative 9; Mismatches 15; Indels 9; Gaps 5;

Db 385 flalaipynhwtkenrigfvmysltffanfgpnattfvvpaefparfirst-chg 443  
 QY 26 FAFPTTP--RWPNEVYIGPPI--TLHFPAHF-QK---FSQPAEIDFKRYVLLCNG 74

## RESULT 6

ID W32302 standard; Protein; 534 AA.  
 AC W32302;  
 DT 27-APR-1998 (first entry)  
 DE Arabidopsis thaliana inorganic phosphate transporter 4.  
 KW Columbia strain; inorganic phosphate transporter 4; IPT4;  
 KW accelerated phosphate uptake; tobacco plant.  
 OS Arabidopsis thaliana.  
 FN W09735984-A1.  
 PD 02-OCT-1997.  
 PF 24-MAR-1997; J00975.  
 PR 25-MAR-1996; JP-094790.  
 PA (MITS-) MITSUI PLANT BIOTECHNOLOGY RES INST.  
 PI Mitsukawa N, Okumura S, Shibata D, Shirano Y;  
 DR WPI: 97-489647/45.  
 DR N-PSDB; T91322, T91325.  
 PT DNA encoding a plant phosphate transporter protein - useful for  
 PT producing e.g. tobacco plants with increased phosphate uptake and  
 PT accelerated growth  
 TS Claim 2; Pages 72-74; 96pp; Japanese.  
 CC The present sequence is Arabidopsis thaliana (Columbia strain)  
 CC inorganic phosphate transporter 4 (IPT4), useful for accelerating  
 CC the uptake of phosphate by plants, e.g. tobacco plants.  
 SQ Sequence 534 AA;

Query Match 10.7%; Score 86; DB 27; Length 534;

Best Local Similarity 42.1%; Pred. No. 1.69e+01;  
 Matches 24; Conservative 9; Mismatches 15; Indels 9; Gaps 5;

Db 388 flalaipynhwtkenrigfvmysltffanfgpnattfvvpaefparfirst-chg 443  
 QY 26 FAFPTTP--RWPNEVYIGPPI--TLHFPAHF-QK---FSQPAEIDFKRYVLLCNG 74

## RESULT 7

ID F70134 standard; Protein; 440 AA.  
 AC F70134;  
 DT 26-APR-1991 (first entry)  
 DE Natural recombinant human lecithin:cholesterol acyltransferase.  
 KW Lecithin:cholesterol acyltransferase; congenital deficiencies;  
 KW renal disease; hepatitis; heart disease polymorphisms.  
 FH Key Location/Qualifiers  
 FT peptide 25..440  
 FT /label= mature lecithin:cholesterol acyltrans-  
 FT ferase  
 FN EP-222591-A.  
 PD 20-MAY-1987.  
 PF 05-NOV-1986; 308624.  
 PR 08-NOV-1985; US-796473.  
 PA (GETH ) GENENTECH INC.  
 PI Baer BW, Drayna DT, Lawn RM, Mclean JW;  
 DR WPI: 87-137586/20.  
 DR N-PSDB; N70191.  
 PT New DNA encoding lecithin:cholesterol acyl-transferase - useful  
 PT in prodn. of enzyme for therapeutic use and in hybridisation  
 PT assays



RESULT 13.

W27083 standard; Protein; 956 AA.

W27083

18-MAR-1998 (first entry)

Human cytomegalovirus helicase.

Human cytomegalovirus helicase; HCMV UL105; open reading frame; orf;

screening; inhibitor; infection; diagnosis; antiviral.

Homo sapiens.

OS GB2711068-A.

17-SEP-1997.

05-MAR-1997; 004575.

04-APR-1996; GB-007118.

14-MAR-1996; US-013389.

(MERT) MERCK & CO INC.

Gottlieb L, Hazuda DJ, Lafemina RL;;

WPI: 97-427906/40.

P-PSDB; W27083.

Novel human cytomegalovirus helicase - used for screening for HCMV

antivirals and in diagnosis of HCMV related diseases

Claim 1; Pages 13-16; 28pp; English.

The present sequence represents a new human cytomegalovirus (HCMV)

helicase which was used in a screening assay for compounds which

inhibit HCMV helicase, preferably with an IC50 of not greater than 200nM.

DNA encoding HCMV helicase, spanning nucleotides 151926 through 154793 of

HCMV AD169 was amplified by PCR primers T95063-4. The amplified product

was digested with BglII and EcoRI and cloned into BglII/EcoRI digested

pB877 vector DNA to yield the plasmid pB877 UL105. The amplified HCMV

UL105 orf sequence and it translational product, helicase, were

determined. The HCMV UL105 orf was recombinant into the baculovirus AcNPV

genome by standard co-transfection protocols. This new HCMV helicase

is also used as a diagnostic tool for diseases resulting from HCMV

infection.

Sequence 956 AA;

Query Match 10.0%; Score 81; DB 26; Length 956;

Best Local Similarity 27.8%; Pred. No. 4.23e+01;

Matches 15; Conservative 15; Mismatches 22; Indels 2; Gaps 2;

Db 723 ykpsllsfeetvmytfrdfrdkfryqlmqrltgrfatplvtyrnrv 776  
 QY 39 YIGFP-ITLLHPAHFOKESPAEIPDKRYVLLC-NGHNPALQOQTHSSRQV 90

## RESULT 14

ID R63788 standard; Protein: 348 AA.  
 AC R63788;  
 DT 21-JUN-1995 (first entry)  
 DE POU domain transcription regulator Skn-li protein.  
 KW POU domain transcription regulator; Skn-li; skin; hair.  
 OS Rattus rattus.  
 PN W09421783-A.  
 PD 29-SEP-1994.  
 PF 14-MAR-1994; U03003.  
 PR 22-MAR-1993; US-035392.  
 PA (REGC) UNIV CALIFORNIA.  
 PI Anderson B, Rosenfeld MG;  
 DR WPI: 94-317004/39.  
 DR Q-PSDB; Q74634.  
 PT DNA encoding POU domain transcription regulators - used to  
 develop prods. for regulation of skin or hair development or  
 control of skin or hair cell proliferation  
 PT Claim 1; Page 32; 69pp; English.  
 PS This alternative POU domain transcription regulator can be considered as  
 the protein expressed by the 11th Oct gene expressed both in humans and  
 rats. Expression of the DNA encoding this protein is specific in skin to  
 identifiable stages of cell proliferation corresponding to  
 development and maturation of the epidermis as well as cyclical hair  
 growth. Controlled expression of these genes may lead to further  
 insight into the regulation of skin development and may be used in  
 the clinical control of skin cell proliferation, e.g. in  
 CC regenerating skin and terminating metastasis of tumor cells.  
 SQ Sequence 348 AA;

Query Match 9.9%; Score 80; DB 12; Length 348;  
 Best Local Similarity 39.3%; Pred. No. 5.07e+01;

Matches 11; Conservative 6; Mismatches 10; Indels 1; Gaps 1;

Db 1 mvsmfslsfkpggfcif-vcflfcfv1 27

QY 1 MVTIPLVFRWFPVCGFRVCKISSPPAF 28

## RESULT 15

ID W95515 standard; Protein: 227 AA.  
 AC W95515;  
 DT 26-MAR-1999 (first entry)  
 DE Fatty acid desaturase homologue from D. discoideum.  
 KW Delta 5 desaturase; recombinant; fatty acid desaturase; FAD; PUFA; oil;  
 KW polyunsaturated fatty acid; linoleic acid; arachidonic acid; linolenic;  
 KW stearidonic acid; eicosapentaenoic acid; malnutrition; feeding formula;  
 KW dietary supplement; prostaglandin; restenosis; angioplasty; inflammation;  
 KW rheumatoid arthritis; psoriasis; osteoporosis; cancer; eczema; AIDS;  
 KW diabetes; cosmetic; animal feed.  
 OS Dictyostelium discoideum.  
 PN W09846764-A1.  
 PD 22-OCT-1998.  
 PF 10-APR-1998; U07421.  
 PR 24-OCT-1997; US-956985.  
 PR 11-APR-1997; US-833610.  
 PR 11-APR-1997; US-834033.  
 PR 11-APR-1997; US-834655.  
 PA (ABBO) ABBOTT LAB.  
 PA (CALJ) CALGENE LLC.  
 PI Chaudhary S, Huang Y, Knutzen D, Leonard AE, Mukerji P,  
 PI Thurmond J;  
 DR WPI: 99-080739/07.  
 DR N-PSDB; X00911.  
 PT Nucleic acid construct able to express fatty acid desaturase in  
 PT plants - useful in human or animal nutrition, as cosmetics and  
 PT therapeutically, e.g. for restenosis, cancer and diabetes

PS Example 18; Page 171-172; 210pp; English.

CC The invention relates to a nucleic acid construct that contains at least  
 CC one of the nucleotide sequences (X00889 to X00891) encoding Mortierella  
 CC alpina delta 6, delta 12 and delta 5 desaturases (W95504 to W95506)  
 CC respectively, coupled to an expression control sequence functional in  
 CC plants. Recombinant plant cells containing at least one DNA encoding a  
 CC M. alpina fatty acid desaturase (FAD), can be used for the production of  
 CC polyunsaturated fatty acid (PUFA). These recombinant cells or plants  
 CC containing them are used to produce oils such as linoleic acid,  
 CC arachidonic acid, gamma-linolenic acid, dihomogamma-linolenic acid,  
 CC used: (i) to treat malnutrition; (ii) in infant feeding formulas; or  
 CC dietary supplements or substitutes, for use in humans or animals; (iii)  
 CC for treating disorders associated with inadequate consumption or  
 CC production of PUFA (or their metabolites such as prostaglandins), e.g.  
 CC restenosis after angioplasty, inflammation, AIDS, rheumatoid arthritis,  
 CC psoriasis, osteoporosis, cancer, eczema, diabetes; (iv) as cosmetics, and  
 CC (v) as animal feeds. Fragments of the DNA are used as probes to isolate  
 CC related coding sequences. Recombinant plants can produce high yields of  
 CC PUFA, since new pathways can be created and unwanted ones suppressed.  
 CC Plants can be engineered to express oils of particular PUFA composition,  
 CC e.g. one similar to that in human milk, and product recovery is simpler  
 CC than with e.g. fish. The present sequence represents a desaturase from  
 CC D. discoideum that shows significant homology to M. alpina delta 5 and  
 CC delta 6 desaturases.  
 SQ Sequence 227 AA;

Query Match 9.8%; Score 79; DB 39; Length 227;  
 Best Local Similarity 29.4%; Pred. No. 6.08e+01;

Matches 10; Conservative 10; Mismatches 13; Indels 1; Gaps 1;

Db 90 iaefvfgwyitinfqshvaedlkxfatperpde 123

QY 4 IPLPLVFRWFPVCGFRVCKISSPPAF-TTPRWPHN 36

Search completed: Wed Dec 22 10:22:33 1999

Job time : 20 secs.

\*\*\*\*\*  
WIREH (TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Dec 22 10:24:18 1999; MasPar time 4.00 Seconds  
309.550 Million cell updates/sec

Tabular output not generated.

Title: >US-09-209-961-9  
Description: (1-104) from US09209961.pep  
Perfect Score: 807  
Sequence: 1 MVTIPPLVFRWPPVCGFRVC.....HSSRQVTPLSLRSRSTFNK 104

Scoring table: PAM 150  
Gap 11

Searched: 122461 seqs, 11912985 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-issued  
1:5A\_COMB 2:5B\_COMB 3:PCT9\_COMB 4:backfiles1

Statistics: Mean 27.055; Variance 110.616; scale 0.245

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				Query		Match		Length		DB		ID		Description		Pred. No.	
Result No.	Score	Match	%	No.	Score	Match	%	No.	Score	Match	%	No.	Score	Match	%	No.	Score
1	88	10.9		1	88	10.9		1	88	10.9		1	88	10.9		1	88
2	86	10.7		2	86	10.7		2	86	10.7		2	86	10.7		2	86
3	81	10.0		3	81	10.0		3	81	10.0		3	81	10.0		3	81
4	80	9.9		4	80	9.9		4	80	9.9		4	80	9.9		4	80
5	80	9.9		5	80	9.9		5	80	9.9		5	80	9.9		5	80
6	79	9.8		6	79	9.8		6	79	9.8		6	79	9.8		6	79
7	79	9.8		7	79	9.8		7	79	9.8		7	79	9.8		7	79
8	79	9.8		8	79	9.8		8	79	9.8		8	79	9.8		8	79
9	79	9.8		9	79	9.8		9	79	9.8		9	79	9.8		9	79
10	79	9.8		10	79	9.8		10	79	9.8		10	79	9.8		10	79
11	79	9.8		11	79	9.8		11	79	9.8		11	79	9.8		11	79
12	79	9.8		12	79	9.8		12	79	9.8		12	79	9.8		12	79
13	79	9.8		13	79	9.8		13	79	9.8		13	79	9.8		13	79
14	78	9.7		14	78	9.7		14	78	9.7		14	78	9.7		14	78
15	77	9.5		15	77	9.5		15	77	9.5		15	77	9.5		15	77
16	77	9.5		16	77	9.5		16	77	9.5		16	77	9.5		16	77
17	77	9.5		17	77	9.5		17	77	9.5		17	77	9.5		17	77
18	77	9.5		18	77	9.5		18	77	9.5		18	77	9.5		18	77
19	77	9.5		19	77	9.5		19	77	9.5		19	77	9.5		19	77
20	76	9.4		20	76	9.4		20	76	9.4		20	76	9.4		20	76
21	76	9.4		21	76	9.4		21	76	9.4		21	76	9.4		21	76
22	76	9.4		22	76	9.4		22	76	9.4		22	76	9.4		22	76
23	75	9.3		23	75	9.3		23	75	9.3		23	75	9.3		23	75

CC TOPOLOGY: unknown  
CC MOLECULE TYPE: protein  
SQ SEQUENCE 775 AA; 87538 MW; 3206870 CN;

Query Match 10.9%; Score 88; DB 1; Length 775;  
Best Local Similarity 39.4%; Pred. No. 6.28e+00;  
Matches 13; Conservative 7; Mismatches 12; Indels 1; Gaps 1;

Db 215 CNEYINGLPP-IQNTNRVPLSLSSRSIOYKR 246  
|| | : | : : | | | | | : :  
QY 72 CNGHNPALQOGTHSSRQVTPLSLRSSTFNK 104

RESULT 2  
ID US-07-603-133B-14 STANDARD; PRT; 775 AA.  
XX  
AC xxxxxx  
XX  
DT  
DE  
DE DE 215 CNEYINGLPP-IQNTNRVPLSLSSRSIOYKR 246  
CC || | : | : : | | | | | : :  
CC QY 72 CNGHNPALQOGTHSSRQVTPLSLRSSTFNK 104  
CC

Sequence 14, Application US/07603133B  
Sequence 14, Application US/07603133B  
Patent No. 5298244  
GENERAL INFORMATION:  
APPLICANT: Redmond, Mark J.  
APPLICANT: Ijaz, Mohammed K.  
APPLICANT: Parker, Michael D.  
TITLE OF INVENTION: ASSEMBLED VIRAL PARTICLES AND THEIR  
TITLE OF INVENTION: USE IN A VACCINE TO ROTAVIRAL DISEASE  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Morrison & Foerster  
STREET: 545 Middlefield Road, Suite 200  
CITY: Menlo Park  
STATE: CA  
COUNTRY: USA  
ZIP: 94025

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/603,133B  
FILING DATE: 19901025  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Robins, Roberta L.  
REGISTRATION NUMBER: 33,208  
REFERENCE/DOCKET NUMBER: 9313-0004.00  
TELEPHONE: (415) 327-7250  
TELEFAX: (415) 327-2951  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 775 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
SQ SEQUENCE 775 AA; 87593 MW; 3180428 CN;

Query Match 10.7%; Score 86; DB 1; Length 775;  
Best Local Similarity 39.4%; Pred. No. 9.07e+00;  
Matches 13; Conservative 6; Mismatches 13; Indels 1; Gaps 1;

Db 215 CNEYINGLPP-IQNTNRVPLSLSSRSIOYKR 246  
|| | : | : : | | | | | : :  
QY 72 CNGHNPALQOGTHSSRQVTPLSLRSSTFNK 104  
CC

RESULT 3  
ID US-08-553-619B-2 STANDARD; PRT; 159 AA.  
XX  
AC xxxxxx  
XX  
DT  
DE  
DE DE 215 CNEYINGLPP-IQNTNRVPLSLSSRSIOYKR 246  
CC || | : | : : | | | | | : :  
CC QY 72 CNGHNPALQOGTHSSRQVTPLSLRSSTFNK 104  
CC

Sequence 2, Application US/08553619B  
Sequence 2, Application US/08553619B  
Patent No. 5919705  
GENERAL INFORMATION:  
APPLICANT: Dehaan, Petrus T.  
TITLE OF INVENTION: Virus Resistant Plants  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 5919705artis Crop Protection  
STREET: 975 California Avenue  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/553,619B  
FILING DATE: December 1, 1995  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Marcus-Wyner, Lynn  
REGISTRATION NUMBER: 34,869  
REFERENCE/DOCKET NUMBER: 137-1082/PCT  
TELEPHONE: 415/354-3588  
TELEFAX: 415/857-1125  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 159 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SQ SEQUENCE 159 AA; 17773 MW; 131586 CN;

Query Match 10.0%; Score 81; DB 2; Length 159;  
Best Local Similarity 28.9%; Pred. No. 2.24e+01;  
Matches 11; Conservative 12; Mismatches 15; Indels 0; Gaps 0;

Db 88 FDTNRRIIEVENQNPITATLDTATRRVDDATVAIRSA 125  
|| | : | : : | | | | | : :  
QY 63 FDKRYRLCNGHNPALQOGTHSSRQVTPLSLRSRS 100  
CC

RESULT 4  
ID US-08-035-392-2 STANDARD; PRT; 348 AA.  
XX  
AC xxxxxx  
XX  
DT  
DE  
DE DE 215 CNEYINGLPP-IQNTNRVPLSLSSRSIOYKR 246  
CC || | : | : : | | | | | : :  
CC QY 72 CNGHNPALQOGTHSSRQVTPLSLRSSTFNK 104  
CC

Sequence 2, Application US/08035392  
Sequence 2, Application US/08035392  
Patent No. 5484732  
GENERAL INFORMATION:  
APPLICANT: Rosenfeld, M. G.  
APPLICANT: Andersen, B.  
TITLE OF INVENTION: TRANSCRIPTION FACTOR FOR THE REGULATION  
TITLE OF INVENTION: OF THE DEVELOPMENT OF SKIN AND HAIR  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:





```
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC FRAGMENT TYPE: internal
CC SEQUENCE 345 AA; 37857 MW; 648414 CN;

Query Match
Best Local Similarity 32.7%; Score 79; DB 2; Length 345;
Matches 16; Conservative 9; Mismatches 21; Indels 3; Gaps 3;

Db 213 ITLPL-FGWAQNVNDKVKCLISQDFGTYYS-TAVAFYIPMSVLMFY 259
QY 2 VTIPPLVFRWFP-VGGRVCKISSPFAFTTPRPHNEVYIGFPITLLHF 49

RESULT 7
ID PCT-US93-10301-9 STANDARD; PRT; 365 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX
Sequence 9, Application PC/TUS9310301
Sequence 9, Application PC/TUS9310301
GENERAL INFORMATION:
APPLICANT: The United States of America, as represented by
APPLICANT: the Secretary of Health and Human Services
TITLE OF INVENTION: THE PCT-65 SEROTONIN RECEPTOR
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: KNOBBE, MARTENS, OLSON AND BEAR
STREET: 620 NEWPORT CENTER DRIVE 16TH FLOOR
CITY: NEWPORT BEACH
STATE: CA
COUNTRY: USA
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION NUMBER: US/08/428,243
FILING DATE: 18-SEP-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/10301
FILING DATE: 27-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: Altman, Daniel E.
REGISTRATION NUMBER: 34,115
REFERENCE/DOCKET NUMBER: NIH046.001VPC
TELEPHONE: 714-760-0404
TELEFAX: 714-760-9502
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 365 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE 365 AA; 41165 MW; 680788 CN;

Query Match
Best Local Similarity 32.7%; Score 79; DB 2; Length 365;
Matches 16; Conservative 9; Mismatches 21; Indels 3; Gaps 3;

Db 135 ITLPL-FGWAQNVNDKVKCLISQDFGTYYS-TAVAFYIPMSVLMFY 181
QY 2 VTIPPLVFRWFP-VGGRVCKISSPFAFTTPRPHNEVYIGFPITLLHF 49

RESULT 9
ID US-08-428-243-7 STANDARD; PRT; 404 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX
Sequence 7, Application US/08428243
Sequence 7, Application US/08428243
GENERAL INFORMATION:
APPLICANT: The United States of America, as represented by
APPLICANT: the Secretary of Health and Human Services
TITLE OF INVENTION: THE PCT-65 SEROTONIN RECEPTOR
NUMBER OF SEQUENCES: 9
```

CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: KNOBBE, MARTENS, OLSON AND BEAR  
CC STREET: 620 NEWPORT CENTER DRIVE 16TH FLOOR  
CC CITY: NEWPORT BEACH  
CC STATE: CA  
CC COUNTRY: USA  
CC ZIP: 92660  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent In Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/428,243  
CC FILING DATE: 18-SEP-1995  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: PCT/US93/10301  
CC FILING DATE: 27-OCT-1993  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Altman, Daniel E.  
CC REGISTRATION NUMBER: 34,115  
CC REFERENCE/DOCKET NUMBER: NIH046.001VPC  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 714-760-0404  
CC FAX: 714-760-9502  
CC INFORMATION FOR SEQ ID NO: 7:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 404 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE: 404 AA; 45041 MW; 847946 CN;  
Query Match 9.8%; Score 79; DB 2; Length 404;  
Best Local Similarity 32.7%; Pred. No. 3.21e+01;  
Matches 16; Conservative 9; Mismatches 21; Indels 3; Gaps 3;  
Db 172 ITLPL-FGWAQNVNDKVLISQDFGTYYS-TAVAFYIPMSVLMFY 218  
QY 2 VTIPPLVRWF-FVCGFRVCKISSPFAFTPRWPHNEVYIGFPTLLHF 49  
RESULT 10  
ID PCT-US93-10301-7 STANDARD; PRT: 404 AA.  
XX XXXXX  
AC  
XX  
DT  
XX  
DE Sequence 7, Application PC/TUS9310301  
XX Sequence 7, Application PC/TUS9310301  
CC GENERAL INFORMATION:  
CC APPLICANT: The United States of America, as represented by  
CC APPLICANT: the Secretary of Health and Human Services  
CC TITLE OF INVENTION: THE PCT-65 SEROTONIN RECEPTOR  
CC NUMBER OF SEQUENCES: 9  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: KNOBBE, MARTENS, OLSON AND BEAR  
CC STREET: 620 NEWPORT CENTER DRIVE 16TH FLOOR  
CC CITY: NEWPORT BEACH  
CC STATE: CA  
CC COUNTRY: USA  
CC ZIP: 92660  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent In Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: PCT/US93/10301  
CC FILING DATE:

CC CLASSIFICATION:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Altman, Daniel E.  
CC REGISTRATION NUMBER: 34,115  
CC REFERENCE/DOCKET NUMBER: NIH046.001VPC  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 714-760-0404  
CC FAX: 714-760-9502  
CC INFORMATION FOR SEQ ID NO: 7:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 404 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE: 404 AA; 45041 MW; 847946 CN;  
Query Match 9.8%; Score 79; DB 3; Length 404;  
Best Local Similarity 32.7%; Pred. No. 3.21e+01;  
Matches 16; Conservative 9; Mismatches 21; Indels 3; Gaps 3;  
Db 172 ITLPL-FGWAQNVNDKVLISQDFGTYYS-TAVAFYIPMSVLMFY 218  
QY 2 VTIPPLVRWF-FVCGFRVCKISSPFAFTPRWPHNEVYIGFPTLLHF 49  
RESULT 11  
ID US-07-996-772A-12 STANDARD; PRT: 422 AA.  
XX XXXXX  
AC  
XX  
DT  
XX  
DE Sequence 12, Application US/07996772A  
XX Sequence 12, Application US/07996772A  
CC Patent No. 5472866  
CC GENERAL INFORMATION:  
CC APPLICANT: Gerald, Christophe  
CC APPLICANT: Hartig, Paul R.  
CC APPLICANT: Branchek, Theresa A.  
CC APPLICANT: Weinshank, Richard L.  
CC TITLE OF INVENTION: DNA ENCODING 5-HT4A SEROTONIN  
CC TITLE OF INVENTION: RECEPTORS AND USES THEREOF  
CC NUMBER OF SEQUENCES: 12  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: COOPER & DUNHAM  
CC STREET: 30 ROCKEFELLER PLAZA  
CC CITY: NEW YORK  
CC STATE: NEW YORK  
CC ZIP: 10112  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent In Release #1.24  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/07/996,772A  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: White, P., John  
CC REGISTRATION NUMBER: 28,678  
CC REFERENCE/DOCKET NUMBER: 42667/JPW/TEP  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (212) 977-9550  
CC FAX: (212) 664-0525  
CC TELETYPE: 422523 COOP UI  
CC INFORMATION FOR SEQ ID NO: 12:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 422 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear



CC STATE: CA  
CC COUNTRY: USA  
CC ZIP: 94608-2916  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent In Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/477,451  
CC FILING DATE: 07-JUN-1995  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: McClung, Barbara G.  
CC REGISTRATION NUMBER: 33,113  
CC REFERENCE/DOCKET NUMBER: 0335,002  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 510-601-2708  
CC INFORMATION FOR SEQ ID NO: 3:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 3174 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 3174 AA; 369950 MW; 52965964 CN;

Query Match 9.7%; Score 78; DB 2; Length 3174;  
Best Local Similarity 36.4%; Pred. No. 3.83e+01;  
Matches 16; Conservative 11; Mismatches 14; Indels 3; Gaps 3;  
Db 237 STPIIVFEKRSKRVALLSKSHNSPLVDLQKGNRATSPLGSLWS 280  
Qy 57 SQPAEIDKRYR-V-LLCNGHONPALQOGTHSSRQVTPL-SLRS 97

RESULT 15  
ID US-08-205-719-4 STANDARD; PRT; 456 AA.  
XX  
AC xxxxxx  
XX  
DT  
XX  
DE  
XX  
XX

Sequence 4, Application US/08205719  
Sequence 4, Application US/08205719  
Patent No. 5705398  
GENERAL INFORMATION:  
APPLICANT: Mintz, D. N.  
APPLICANT: Tobias, P. S.  
APPLICANT: Ulevitch, R. J.  
TITLE OF INVENTION: SYSTEM AND METHODS FOR IDENTIFYING  
TITLE OF INVENTION: THERAPEUTIC INHIBITORS OF LPS-MEDIATED SEPSIS  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: The Scripps Research Institute, Office of  
ADDRESSEE: Patent Counsel  
STREET: 10666 No. 5705398th Torrey Pines Road, TPC 8  
CITY: La Jolla  
STATE: CA  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/205,719  
FILING DATE: 02-MAR-1994  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:

CC NAME: Fitting, Thomas  
CC REGISTRATION NUMBER: 34,163  
CC REFERENCE/DOCKET NUMBER: SCRI201P; TSRI324.0  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 619-554-2937  
CC TELEFAX: 619-554-6312  
CC INFORMATION FOR SEQ ID NO: 4:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 456 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 456 AA; 51494 MW; 1089916 CN;

Query Match 9.5%; Score 77; DB 1; Length 456;  
Best Local Similarity 31.5%; Pred. No. 4.57e+01;  
Matches 23; Conservative 15; Mismatches 31; Indels 4; Gaps 4;  
Db 257 SDYVNTASLAYHKSGYWNFSITDAMPADL-NIRRTKSF-RPFVPLLANLYPNMNL 314  
Qy 24 SPFAFTTPRPW-HNEVYIGFPITLLHFPAPHFQKFSQPAEIDKRYRVLLCNGHONPALQ-81  
Db 315 QGTVNSEQLVNL 327  
Qy 82 OGTHSSRQVTPLS 94

Search completed: Wed Dec 22 10:24:26 1999  
Job time : 8 secs.

Result No.	Query %		DB	ID	Description	Pred. No.
	Score	Match Length				
1	104	12.9	2	S70114	probable membrane pro	5.39e-03
2	92	11.4	2	S74900	histidinol-phosphate	2.94e-01
3	91	11.3	2	T01397	LTR transposon polypr	4.05e-01
4	90	11.2	4	S52460	hypothetical protein	5.57e-01
5	89	11.0	2	D58933	hypothetical protein	7.64e-01
6	89	11.0	700	T20951	probable UvrD - Mycob	7.64e-01
7	89	11.0	793	JC5539	Smoothened protein pr	7.64e-01
8	88	10.9	280	1 VPRXDS	outer capsid protein	1.05e+00
9	88	10.9	469	2 S71813	crty protein - Rhodob	1.05e+00
10	88	10.9	775	1 VPRXHK	outer layer protein V	1.05e+00
11	88	10.9	775	1 VPRXW5	outer layer protein V	1.05e+00
12	87	10.8	631	2 T02472	hypothetical protein	1.43e+00
13	86	10.7	421	2 S35301	rfdH protein - Yersin	1.94e+00
14	86	10.7	775	1 VPRXRM	outer layer protein V	1.94e+00
15	86	10.7	775	1 VPRXWL	outer layer protein V	1.94e+00
16	86	10.7	775	1 VPRXW4	outer layer protein V	1.94e+00
17	86	10.7	775	2 S52165	outer capsid protein	1.94e+00
18	86	10.7	775	1 VPRXPG	outer layer protein V	1.94e+00
19	85	10.5	147	2 S00775	hypothetical protein	2.64e+00
20	85	10.5	234	2 S70539	response regulator ci	2.64e+00
21	85	10.5	280	1 VPRXWA	outer capsid protein	2.64e+00
22	85	10.5	280	1 VPRX16	outer capsid protein	2.64e+00
23	85	10.5	288	2 S75502	hypothetical protein	2.64e+00

```

ACCESSIONS S74900      nucleic acid sequence not shown; translation not shown
REFERENCE   S74322      #molecule_type DNA
AUTHOR     Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
            Nakamura, Y.; Miyajima, N.; Hiroseawa, M.; Sugiyura, M.;
            Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
            Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo,
            S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
            Yasuda, M.; Tabata, S.
#journal   DNA Res. (1996) 3:109-136
#title     Sequence analysis of the genome of the unicellular
            cyanobacterium Synecocystis sp. PC6803. II. Sequence
            determination of the entire genome and assignment of
            potential protein-coding regions.
#cross-references UID:97061201
#accession S74900      #label KAN
                        1-367 #residues
                        EMBL:D90909; GB:AB001339; NID:g1652844; PID:d1018594;
                        PID:g1652943
#note      the nucleotide sequence was submitted to the EMBL Data
            Library, June 1996

GENETICS   hisC-2
#gene      #length 367 #molecular-weight 40686 #checksum 1659
SUMMARY    Query Match          11.4%; Score 92; DB 2; Length 367;
            Best Local Similarity 37.5%; Pred. No. 2,94e-01;
            Matches           9; Conservative 7; Mismatches 7; Indels 1; Gaps 1;

Db         297 FQVCPSDANFVFAPRKNWAADLY 320
            !::: : !:::!! :: !!
QY         17 FRVCKISSPFAFTPRW-PHNEVY 39

```

```

3
RESULT      3
ENTRY      T01397
TITLE      LTR transposon polyprotein homolog T419.16 - Arabidopsis thaliana
ORGANISM   Arabidopsis thaliana
#formal_name Arabidopsis thaliana
#common_name mouse-ear cress
DATE       12-Feb-1999
#sequence_revision 12-Feb-1999
#text_change 24-Mar-1999
ACCESSION  T01397
REFERENCE  Z14314
#authors   Patnell, L.D.; Gnoj, L.; de la Bastide, M.; Hameed, A.; Habermann, K.; Schutz, K.; Huang, E.; Gottesman, T.; Dedhia, N.N.; McCombie, W.R.
#submision submitted to the EMBL Data Library, May 1998
#description Genomic sequence of BAC T419 from Arabidopsis thaliana, chromosome IV, near 16.6 cM.
#accession T01397
#status     translated from GB/EMBL/DBJ
#molecule_type DNA
#residues   1-1456
#label      PAR
#cross-references EMBL:AF069442; NID:g3242970; PID:g3924609
#experimental_source cultivar Columbia
GENETICS
#map_position 4
#introns      129/1
#note         T419.16
SUMMARY      #length 1456 #molecular-weight 162635 #checksum 5916
Query Match 11.3% Score 91; DB 2; Length 1456;
Best Local Similarity 39.5%; Pred.No. 4.05e-01;
Matches 15; Conservative 11; Mismatches 7; Indels 5; Gaps 5;
Db 637 VYILNRLPTLLQLQSPFKLFGPPN-YEK-LKVFQC 672
QY 38 VY-IG-FPTLLHFAHFQK-FSQPAEFDKRYRLVC 72
RESULT      4

```

```

ENTRY          S52460      #type complete
TITLE          hypothetical protein I09 - Coxiella burnetii
ORGANISM       COXIELLA BURNETII
DATE           08-May-1995 #sequence_revision 21-Jul-1995 #text_change
               09-Sep-1997
ACCESSIONS     S52460
REFERENCE      S52460
#authors       Willems, H.; Thiele, D.; Valkova, D.
#submission    submitted to the EMBL Data Library, February 1995
#accession     S52460
#status        preliminary
##molecule_type DNA
##residues     1-82 ##label WIL
##cross-references EMBL:X84722; NID:g682754; PID:g682755
SUMMARY        #length 82 #molecular-weight 9545 #checksum 1991

Query Match   11.2%; Score 90; DB 2; Length 82;
Best Local Similarity 20.8%; Pred. No. 5.57e-01;
Matches 10; Conservative 18; Mismatches 19; Indels 1; Gaps 1;

Db 28 LLLVPSLVLFMWFPFISSIDAYLIYQYGILATWNLCDVLTYYLASL 75
Qy 1 MYTIPPLV-FRFPVGGRVKCKISSPFAFTTPRPHPNEVIIGFPIILL 47
:::||||| ::::: :|| ::::: || ::::: || ::::: || ::::: ||
#cross-references GB:D89961; NID:g4115781; PID:g4115800
#accession     D58933
#status        preliminary
##molecule_type DNA
#residues      1-267 ##label ARN
#cross-references GB:D89961; NID:g4115781; PID:g4115800
GENETICS       ORF267
#gene          mitochondrion
#genome        mitochondrion
KEYWORDS        length 267 #molecular-weight 32317 #checksum 5048
SUMMARY

Query Match   11.0%; Score 89; DB 2; Length 267;
Best Local Similarity 37.1%; Pred. No. 7.64e-01;
Matches 13; Conservative 10; Mismatches 9; Indels 3; Gaps 3;

Db 167 NVYI-FPISLLVFSEKRYTKL-AKNY-RRYLLLL 198
Qy 37 EYVIGFPIILHFFPAHFQKFSQPAPAEIFDKRYRVLL 71
:::||||| ::::: :|| ::::: || ::::: || ::::: || ::::: ||
#cross-references GB:D89961; NID:g4115781; PID:g4115800
#accession     D70951
#status        type complete
PROBLEMA URVD - Mycobacterium tuberculosis (strain H37RV)
ORGANISM       MYCOBACTERIUM TUBERCULOSIS
DATE           17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
               17-Jul-1998

```

D70951  
 ACCESSIONS  
 REFERENCE  
 A70500  
 authors  
 Cole, S. T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eglmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.;

```

Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
Skellton, S.; Squares, S.; Squires, R.; Sulston, J.E.;
Taylor, K.; Whitehead, S.; Barrell, B.G.
Nature (1998) 393:537-544
#journal
#title
Deciphering the biology of Mycobacterium tuberculosis from
the complete genome sequence.
#cross-references MUID:98295987
#accession D70951
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-700 #label COL
#cross-references GB:AL021646; GB:AL123456; NID:G3242278; PID:e1248789;
PID:G2827608
#experimental_source strain H37Rv
GENETICS
#gene
#summary
#length 700 #molecular-weight 75603 #checksum 9430
Query Match 11.0%; Score 89; DB 2: Length 700;
Best Local Similarity 34.1%; Pred. No. 7.64e-01;
Matches 15; Conservative 12; Mismatches 16; Indels 1; Gaps 1;
Db 202 LDFDOLLHTAAIENDAVAEEFQDRYCFVDEYQDVTPLQ 245
QY 40 IGFPITLLHFAHFKQSPAEIFDKRYVLLCNGHQN-PALQ 82
RESULT 7
ENTRY #type complete
TITLE Smoothed protein precursor - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Sep-1997 #sequence_revision 05-Sep-1997 #text_change
17-Mar-1999
ACCESSION JC5539; PC4476
REFERENCE JC5539
#authors Akiyama, H.; Shigeno, C.; Hiraki, Y.; Shukunami, C.; Kohno,
H.; Akagi, M.; Konishi, J.; Nakamura, T.
#journal Biochem. Biophys. Res. Commun. (1997) 235:142-147
#title Cloning of a mouse Smoothed cDNA and expression patterns of
hedgehog signalling molecules during chondrogenesis and
cartilage differentiation in clonal mouse EC cells, ATDC5.
#cross-references MUID:97339452
#accession JC5539
#molecule_type mRNA
#residues 1-793 #label AK1
#accession PC4476
#molecule_type protein
#residues 528-533:539-545; 600-605 #label AK2
#experimental_source ATDC5 cell
#comment This protein is used in the conserved targets in Hedgehog
signalling pathway, together with patched and Gli. These protein
are responsible for the skeletal abnormalities in Gorlin and
Greig syndromes.
FEATURE
1-32 #domain signal sequence #status predicted #label SIG
SUMMARY #length 793 #molecular-weight 87299; #checksum 8609
Query Match 11.0%; Score 89; DB 2: Length 793;
Best Local Similarity 27.9%; Pred. No. 7.64e-01;
Matches 12; Conservative 13; Mismatches 16; Indels 2; Gaps 2;
Db 461 FLAFGVLTITSFHYDFNOAE-WERSFRDYVLCQAVNTIGL 502
QY 39 YIGFPITLLHFAHFKQSPAEIFDKRYV-LLCNGHQN-PAL 80
RESULT 8
ENTRY #type complete
TITLE outer capsid protein VP8 - human rotavirus A (serotype 2
strain DS1)

```

```

ORGANISM #formal_name human rotavirus A
DATE 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change
30-Sep-1993
ACCESSION C25904
REFERENCE A94126
#authors Gorziglia, M.; Hoshino, Y.; Buckler-White, A.; Blumentals,
I.; Glass, R.; Flores, J.; Kapikian, A.Z.; Chanock, R.M.
#journal Proc. Natl. Acad. Sci. U.S.A. (1986) 83:7039-7043
#title Conservation of amino acid sequence of VP8 and cleavage
region of 84-kDa outer capsid protein among rotaviruses
recovered from asymptomatic neonatal infection.
#cross-references MUID:86313706
#accession C25904
#molecule_type genomic RNA
#residues 1-280 #label GOR
#comment VP8 is one of two trypsin cleavage products of VP3; the other
product is VP5.
GENETICS
#map_position segment 4
#classification #superfamily rotavirus outer layer protein VP3
#keywords glycoprotein; outer capsid protein
#feature 56,132,150,193 #binding_site carbohydrate (Asn) (covalent) #status
predicted
#summary #length 280 #molecular-weight 32010 #checksum 1604
Query Match 10.9%; Score 88; DB 1: Length 280;
Best Local Similarity 39.4%; Pred. No. 1.05e+00;
Matches 13; Conservative 7; Mismatches 12; Indels 1; Gaps 1;
Db 215 CNEYINGLPP-IQNRNVVPLSLSSRSIQYKR 246
QY 72 CNGHNPALQQTGTHSSRQVTPLSLRSSTENK 104
RESULT 5
ENTRY #type complete
TITLE crtJ protein - Rhodobacter capsulatus
ORGANISM #formal_name Rhodobacter capsulatus
DATE 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
30-Jun-1992
ACCESSION S17813
REFERENCE S17803
#authors Burke, D.H.; Alberti, M.; Armstrong, G.A.; Hearst, J.E.
#submission submitted to the EMBL Data Library, November 1991
#description The complete nucleotide sequence of the 45 kb photosynthesis
gene cluster of Rhodobacter capsulatus.
#accession S17813
#molecule_type DNA
#residues 1-469 #label EMB
#cross-references EMBL:211165
GENETICS crtJ
SUMMARY #length 469 #molecular-weight 51331 #checksum 8204
Query Match 10.9%; Score 88; DB 2: Length 469;
Best Local Similarity 35.0%; Pred. No. 1.05e+00;
Matches 14; Conservative 10; Mismatches 13; Indels 3; Gaps 3;
Db 147 YEAQREI-ETRYRVLL-EAHPAPLLIVSMSTGR-IADLNL 183
QY 56 FSPQAEIFDKRYVLLCNGHQN-PALQQTGTHSSRQVTPLSL 95
RESULT 10
ENTRY #type complete
TITLE outer layer protein VP3 - human rotavirus A (strain KU)
ALTERNATE_NAMES glycoprotein VP3; hemagglutinin; outer capsid protein VP3
ORGANISM #formal_name human rotavirus A
DATE 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change
21-Nov-1997
ACCESSION A28844
REFERENCE A28844

```

```

#authors      Taniguchi, K.; Maloy, W.L.; Nishikawa, K.; Green, K.Y.;
               Hoshino, Y.; Urasawa, S.; Kapikian, A.Z.; Chanock, R.M.;
               Gorziglia, M.
#journal      J. Virol. (1988) 62:2421-2426
#title       Identification of cross-reactive and serotype 2-specific
               neutralization epitopes on VP3 of human rotavirus.
#cross-references MUID:88230603
#accession   A28844
               ##molecule_type mRNA
               ##residues 1-775 ##label TAN
               ##cross-references GB:M21014; NID:g333852; PID:g333853
GENETICS
#map_position segment 4
CLASSIFICATION #superfamily rotavirus outer layer protein VP3
KEYWORDS       glycoprotein; hemagglutinin; outer capsid protein
FEATURE
32,56,97,132,324, #binding_site carbohydrate (Asn) (covalent) #status
583,589,592,599   predicted
SUMMARY        #length 775 #molecular-weight 87538 #checksum 4552
               Query Match 10.9%; Score 88; DB 1; Length 775;
               Best Local Similarity 39.4%; Pred. No. 1.05e+00;
               Matches 13; Conservative 7; Mismatches 12; Indels 1; Gaps 1;
Db 215 CNEYIINGLPP-IQNTNRVVPPLSLRSIOYKR 246
|| | | | | | | | | | | | | | | | | | | | |
QY 72 CNGHQPALQOQGHSSRQVTPLSLRSSSTFNK 104

RESULT 11      VPXRW5      #type complete
ENTRY          outer layer protein VP3 - human rotavirus A (serotype 3
TITLE          strain P)
CONTAINS       hemagglutinin
OUTER CAPSID  outer capsid protein VP5; outer capsid protein VP8
ORGANISM       #formal_name human rotavirus A
DATE          30-Jun-1990 #sequence_revision 30-Jun-1990 #text_change
               01-Mar-1996
ACCESSIONS    C28839
REFERENCE     A28839
#authors      Gorziglia, M.; Green, K.; Nishikawa, K.; Taniguchi, K.;
               Jones, R.; Kapikian, A.Z.; Chanock, R.M.
#journal      J. Virol. (1988) 62:2978-2984
#title       Sequence of the fourth gene of human rotaviruses recovered
               from asymptomatic or symptomatic infections.
#cross-references MUID:88275070
#accession   C28839
               ##molecule_type genomic RNA
               ##residues 1-775 ##label GOR
GENETICS
#map_position segment 4
CLASSIFICATION #superfamily rotavirus outer layer protein VP3
KEYWORDS       capsid protein; coat protein; glycoprotein; hemagglutinin
FEATURE
1-240          #product outer capsid protein VP8 #status predicted
               #label VP8\
241-246        #region cleavage processing #status predicted\
247-775        #product outer capsid protein VP5 #status predicted
               #label VP5\
32,56,97,129,132,
195,324,583,589,
592,599        #binding_site carbohydrate (Asn) (covalent) #status
               predicted
SUMMARY        #length 775 #molecular-weight 87687 #checksum 7244
               Query Match 10.9%; Score 88; DB 1; Length 775;
               Best Local Similarity 39.4%; Pred. No. 1.05e+00;
               Matches 13; Conservative 7; Mismatches 12; Indels 1; Gaps 1;
Db 215 CNEYIINGLPP-IQNTNRVVPPLSLRSIOYKR 246
|| | | | | | | | | | | | | | | | | | | | |
QY 72 CNGHQPALQOQGHSSRQVTPLSLRSSSTFNK 104

```

```

RESULT 12
ENTRY      T02472      #type complete
TITLE      hypothetical protein F4118.27 - Arabidopsis thaliana
ORGANISM   #formal_name Arabidopsis thaliana #common_name mouse-ear
               cress
DATE       05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change
ACCESSIONS T02472
REFERENCE   Z14674
#authors    Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.;
               Brandon, R.C.; Sykes, S.M.; Kaul, S.; Mason, T.M.;
               Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter,
               J.C.
#submission submitted to the EMBL Data Library, August 1998
#description Arabidopsis thaliana chromosome II BAC F4118 genomic
               sequence.
#accession  T02472
               ##status preliminary; translated from GB/EMBL/DBJ
               ##molecule_type DNA
               ##residues 1-631 ##label ROU
               ##cross-references EMBL:AC004665; NID:g3386593; PID:g3386620
GENETICS
#map_position 2
#introns      218/3; 289/1; 389/2; 523/2
#note        F4118.27
SUMMARY      #length 631 #molecular-weight 72101 #checksum 5374
               Query Match 10.8%; Score 87; DB 2; Length 631;
               Best Local Similarity 50.0%; Pred. No. 1.43e+00;
               Matches 8; Conservative 6; Mismatches 1; Indels 1; Gaps 1;
Db 130 CRIPAPYGYKTPFRWP 145
|| | | | | | | | | | | | | | | | | | | | |
QY 20 CKISSPFAFTTP-RWP 34

RESULT 13
ENTRY      S35301      #type complete
TITLE      rfbH protein - Yersinia enterocolitica
ORGANISM   #formal_name Yersinia enterocolitica
DATE       13-Jan-1995 #sequence_revision 06-Jun-1997 #text_change
               17-Mar-1999
ACCESSIONS  S35301; S28584
REFERENCE   S35292
#authors    Zhang, L.; Al-Hendy, A.; Toivanen, P.; Skurnik, M.
#journal    Mol. Microbiol. (1993) 9:309-321
#title      Genetic organization and sequence of the rfb gene cluster of
               Yersinia enterocolitica serotype O:3; similarities to the
               GDP-L-rhamnose biosynthesis pathway of Salmonella and to
               the bacterial polysaccharide transport systems.
#cross-references MUID:94018626
#accession  S35301
               ##status nucleic acid sequence not shown; translation not shown
               ##molecule_type DNA
               ##residues 1-421 ##label ZHA
               ##cross-references EMBL:Z18920; NID:g48582; PID:g48592
               ##experimental_source serotype O:3
               ##note the nucleotide sequence was submitted to the EMBL Data
                       Library, December 1992
GENETICS
#gene       rfbH
SUMMARY     #length 421 #molecular-weight 47718 #checksum 3631
               Query Match 10.7%; Score 86; DB 2; Length 421;
               Best Local Similarity 35.0%; Pred. No. 1.94e+00;
               Matches 14; Conservative 7; Mismatches 17; Indels 2; Gaps 2;
Db 86 TTFIWFVPMWS-N-VWGGGHTLFRFANHFANFNRIIY 123
|| | | | | | | | | | | | | | | | | | | | |
QY 24 SPFAFTTPRPHNEVYIGFPIITLLHFPAPKFSQPAEIF 63

```



```

1-240...      #product outer capsid protein VP8 #status predicted
               #label VP8\
241-246      #region cleavage processing #status predicted\
247-775      #product outer capsid protein VP5 #status predicted
               #label VP5\

  32,56,97,132,150,
  195,324,583,589,
  599
#binding_site carbohydrate (Asn) (covalent) #status
predicted
SUMMARY      #length 775 #molecular-weight 87650 #checksum 6192

Query Match      10.7%; Score 86; DB 1; Length 775;
Best Local Similarity 39.4%; Pred. No. 1.94e+00;
Matches 13; Conservative 6; Mismatches 13; Indels 1; Gaps 1;

Db      215 CNEYINNGLPP-IONTRNVVPLSLSSRSIQYRR 246
      ||| | | | | | | | | | | | | | | |
Q7      72 CNGCHQNPALQOQTHSSRQVTPLSLRSRSTFNK 104

Search completed: Wed Dec 22 10:23:03 1999
Job time : 12 secs.

```

\*\*\*\*\*  
 W A S R E L H  
 \*\*\*\*\* (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
 Run on: Wed Dec 22 10:23:20 1999; MasPar time 5.92 Seconds  
 Tabular output not generated.  
 --496.197 Million cell updates/sec

Title: >US-09-209-961-9  
 Description: (1-104) from US09209961.pep  
 Perfect Score: 807  
 Sequence: 1 MVTIPPLVFRWFPVCGFRVC.....HSSQVTPLSLRSSSTFNK 104

Scoring table: PAM 150  
 Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%

Database: Listing first 45 summaries

Database: swiss-prot37  
 1:swissprot

Statistics: Mean 40.593; Variance 66.688; scale 0.609

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	92	11.4	367	1 H188_SVNY3	HISTIDINOL-PHOSPHATE A	7.25e-02
2	88	10.9	469	1 CRTJ_RHOCA	OUTER CAPSID PROTEIN V	2.97e-01
3	88	10.9	775	1 VP4_ROTTP	OUTER CAPSID PROTEIN V	2.97e-01
4	88	10.9	775	1 VP4_ROTTP	OUTER CAPSID PROTEIN V	2.97e-01
5	87	10.8	524	1 CP51_UNCNE	CYTOTOXIC P450 51 (EC 4.2.0e-01	5.92e-01
6	86	10.7	775	1 VP4_ROTTP	OUTER CAPSID PROTEIN V	5.92e-01
7	86	10.7	775	1 VP4_ROTTP	OUTER CAPSID PROTEIN V	5.92e-01
8	86	10.7	775	1 VP4_ROTTP	OUTER CAPSID PROTEIN V	5.92e-01
9	85	10.5	147	1 YJVL_YEAST	VERY HYPOPHOSPHATE 16.8	8.32e-01
10	85	10.5	234	1 CITB_KLEPN	TRANSFERRIN HOMOLOG ( 8.32e-01	8.32e-01
11	85	10.5	355	1 NIFS_LACDE	NIFS PROTEIN HOMOLOG ( 8.32e-01	8.32e-01
12	85	10.5	440	1 LCAT_RAT	PHOSPHATIDYLCHOLINE-ST	8.32e-01
13	85	10.5	775	1 VP4_ROTTP	OUTER CAPSID PROTEIN V	8.32e-01
14	85	10.5	775	1 VP4_ROTTP	OUTER CAPSID PROTEIN V	8.32e-01
15	85	10.5	775	1 VP4_ROTTP	OUTER CAPSID PROTEIN V	8.32e-01
16	84	10.4	153	1 ELAA_ECOLI	GLUTATHIONE S-TRANSFER	1.17e+00
17	84	10.4	201	1 GT_ECOLI	GLUTATHIONE S-TRANSFER	1.17e+00
18	84	10.4	440	1 LCAT_HUMAN	PHOSPHATIDYLCHOLINE-ST	1.17e+00
19	84	10.4	457	1 UVRD_MCLE	PUTATIVE DNA HELICASE	1.17e+00
20	82	10.2	352	1 YDGO_ECOLI	HYPOTHETICAL 38.1 KD P	2.27e+00
21	82	10.2	438	1 LCAT_MOUSE	PHOSPHATIDYLCHOLINE-ST	2.27e+00
22	82	10.2	440	1 LCAT_PAPAN	PHOSPHATIDYLCHOLINE-ST	2.27e+00
23	82	10.2	775	1 VP4_ROTTP	OUTER CAPSID PROTEIN V	2.27e+00

24	81	10.0	246	1 TF2D_TETH	TRANSCRIPTION INITIATI	3.15e+00
25	31	10.0	425	1 5H7_XENLA	5-HYDROXYTRYPTAMINE 7	3.15e+00
26	81	10.0	956	1 HELI_HCHVA	PROBABLE HELICASE.	3.15e+00
27	80	9.9	78	1 VE5_HPV70	PROBABLE E5 PROTEIN.	4.35e+00
28	80	9.9	326	1 SCRR_PEDPE	SUCROSE OPERON REPRESS	4.35e+00
29	80	9.9	348	1 SKIL_RAT	TRANSCRIPTION FACTOR S	4.35e+00
30	80	9.9	379	1 HYPD_ALCEU	HYDROGENASE EXPRESSION	4.35e+00
31	80	9.9	463	1 GUTA_BACST	PUTATIVE SUGAR TRANSPO	4.35e+00
32	80	9.9	717	1 AMYM_BAGST	MALTOGENIC ALPHA-AMTIA	4.35e+00
33	80	9.9	740	1 POLB_MAIZE	RETROVIRUS-RELATED POL	4.35e+00
34	80	9.9	1325	1 Y309_MYCPN	HYPOTHETICAL LIPOPROTE	4.35e+00
35	79	9.8	183	1 YQHA_HAEIN	HYPOTHETICAL PROTEIN H	6.00e+00
36	79	9.8	305	1 HOG1_HORVO	GAMMA-HORDEIN 1 PRECUR	6.00e+00
37	79	9.8	358	1 ISL3_BRARE	INSULIN GENE ENHANCER	6.00e+00
38	79	9.8	429	1 YKC9_CABEL	HYPOTHETICAL 47.8 KD P	6.00e+00
39	79	9.8	444	1 TIG_MYCGE	TRIGGER FACTOR (TF).	6.00e+00
40	79	9.8	445	1 5H7_HUMAN	5-HYDROXYTRYPTAMINE 7	6.00e+00
41	79	9.8	446	1 5H7_CAVPO	5-HYDROXYTRYPTAMINE 7	6.00e+00
42	79	9.8	448	1 5H7_RAT	5-HYDROXYTRYPTAMINE 7	6.00e+00
43	79	9.8	448	1 5H7_MOUSE	5-HYDROXYTRYPTAMINE 7	6.00e+00
44	79	9.8	451	1 IRF4_HUMAN	INTERFERON REGULATORY	6.00e+00
45	79	9.8	883	1 APCE_CYAPA	PHYCOBILISOME LINKER P	6.00e+00

## ALIGNMENTS

RESULT 1.  
 ID HIS8\_SVNY3 STANDARD; PRT; 367 AA.  
 AC P73807;  
 DT 15-JUL-1998 (REL. 36, CREATED)  
 DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)  
 DE HISTIDINOL-PHOSPHATE AMINOTRANSFERASE (EC 2.6.1.9) (IMIDAZOLE ACETOL-PHOSPHATE TRANSAMINASE)  
 GN HISC OR SLL1958.  
 OS SYNECHOCYSTIS SP. (STRAIN PCC 6803).  
 OC BACTERIA; CYANOBACTERIA; CHROCOCCALES; SYNECHOCYSTIS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 97061201.  
 RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y., MIYAJIMA M., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T., HOSOUCHI T., MATSUO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S., SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M., TABATA S.,  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions."  
 RL DNA RES. 3:109-136(1996).  
 CC 1- CAVALYTIC ACTIVITY: L-HISTIDINOL-PHOSPHATE + 2-OXOGLUTARATE -> 3-(IMIDAZOL-4-YL)-2-OXOPROPYL PHOSPHATE + GLUTAMATE.  
 CC 1- COFACTOR: PYRIDOXAL PHOSPHATE.  
 CC 1- PATHWAY: EIGHTH STEP IN HISTIDINE BIOSYNTHETIC PATHWAY.  
 CC 1- SIMILARITY: BELONGS TO CLASS-II OF PYRIDOXAL-PHOSPHATE-DEPENDENT AMINOTRANSFERASES.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; D90909; G1652943;  
 CC PROSITE; P500599; AA\_TRANSFER\_CLASS\_2; 1.  
 CC PFM; PF00222; aminotran\_2; 1  
 CC HISTIDINE BIOSYNTHESIS; TRANSFERASE; AMINOTRANSFERASE;  
 CC PYRIDOXAL PHOSPHATE.  
 CC BINDING 225  
 CC SEQUENCE 367 AA; 40686 MW; 05AA51CC CRC32;  
 Query Match 11.4%; Score 92; DB 1; Length 367;

GORZIGLIA M., HOSHINO Y., BUCKLER-WHITE A., BLUMENTALS I., GLASS R.  
RA  
FLORES J., KAPIKIAN A.Z., CHANOCK R.M.;  
RT "Conservation of amino acid sequence of VP8 and cleavage region of

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement>, or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC  
DR EMBL; M21014; G333853;  
DR PIR; A28844; VPXRHK.

DR PFAM: PF00426; VP4; 1.  
KW COAT PROTEIN; GLYCOPROTEIN.  
FT CHAIN 1 775  
FT CHAIN 1 240  
FT CHAIN 247 775  
FT CARBOHYD 32 32  
FT CARBOHYD 56 56  
FT CARBOHYD 97 97  
FT CARBOHYD 132 132  
FT CARBOHYD 324 324  
FT CARBOHYD 583 583  
FT CARBOHYD 589 589  
FT CARBOHYD 592 592  
FT CARBOHYD 599 599  
SQ SEQUENCE 775 AA; 87538 MW; AA26C921 CRC32;  
Query Match 10.9%; Score 88; DB 1; Length 775;  
Best Local Similarity 39.4%; Pred. No. 2.97e-01;  
Matches 13; Conservative 7; Mismatches 12; Indels 1; Gaps 1;  
Db 215 CNEYINNGLPP-IQONTRNVVPLSLSSRSIOYKR 246  
QY 72 CNGHONPALQOQTHSSRQVTPLSRSSTFNK 104  
RESULT 5  
ID CP51\_UNCNE STANDARD; PRT; 524 AA.  
AC O1442; O14422;  
DT 15-DEC-1998 (REL. 37, CREATED)  
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)  
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
DE CYTOCHROME P450 51 (EC 1.14.14.1) (CYPL1) (P450-L1A1) (P450-14DM)  
DE ALPHA DEMETHYLASE (EBURICOL 14-ALPHA-DEMETHYLASE) (P450-14DM)  
GN CYP51.  
OS UNCINULA NECATOR (GRAPE POWDERY MILDEW).  
OC EURKOTOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; ERYSPHALES;  
OC ERYSPHACEAE; UNCINULA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN-FPE11;  
RX MEDLINE; 97444284.  
RA DELIE C., LAIGRET F., CORIO-COSTET M.-F.;  
RT "Cloning and sequence analysis of the eburicol 14alpha-demethylase  
gene of the obligate biotrophic grape powdery mildew fungus.";  
RT GENE 195:29-33(1997).  
RN [2]  
RP MOLECULAR BASIS OF FUNGICIDE RESISTANCE.  
RC STRAIN-PAG11;  
RX MEDLINE; 97394921.  
RA DELIE C., LAIGRET F., CORIO-COSTET M.-F.;  
RT "A mutation in the 14 alpha-demethylase gene of Uncinula necator that  
correlates with resistance to a sterol biosynthesis inhibitor.";  
RL APPL. ENVIRON. MICROBIOL. 63:2966-2970(1997).  
CC -1- FUNCTION: CATALYSES C14-DEMETHYLATION OF LANOSTEROL WHICH IS  
CRITICAL FOR ERGOSTEROL BIOSYNTHESIS. IT TRANSFORMS LANOSTEROL  
INTO 4,4'-DIMETHYL CHOLESTA-8,14,24-TRIENE-3-BETA-OL  
(BY SIMILARITY).  
CC -1- PATHWAY: ERGOSTEROL BIOSYNTHESIS.  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
DB EMBL; U72657; G2406574;  
DB EMBL; U72658; G2406576;  
DB EMBL; U83840; G2327023;  
DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
DR PFAM; PF00067; P450; 1.

KW ELECTRON TRANSPORT; OXIDOREDUCTASE; MONOOXYGENASE; MEMBRANE; HEME;  
STEROL BIOSYNTHESIS.  
FT BINDING 469 469  
FT VARIANT 136 136  
FT Y -> F (IN STRAINS RESISTANT TO  
TRIADIMENOL; A STEROL DEMETHYLATION-  
INHIBITING FUNGICIDE).  
SQ SEQUENCE 524 AA; 59855 MW; 10D3D2A2 CRC32;  
Query Match 10.8%; Score 87; DB 1; Length 524;  
Best Local Similarity 41.7%; Pred. No. 4.20e-01;  
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;  
Db 52 PPVVFHWFPIIGTSYGDIPYKF 75  
QY 5 PPIVERWEPVCGFRVCKISSPFAF 28  
RESULT 6  
ID VP4\_ROTGP STANDARD; PRT; 775 AA.  
AC P23045;  
DT 01-NOV-1991 (REL. 20, CREATED)  
DT 01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)  
DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)  
DE OUTER CAPSID PROTEIN VP4 (HEMAGGLUTININ) (OUTER LAYER PROTEIN VP4)  
DE [CONTAINS: OUTER CAPSID PROTEINS VP5 AND VP8].  
GN S4.  
OS PORCINE ROTAVIRUS (STRAIN GOTTFRIED).  
OC VIRUSES; DSRNA VIRUSES; REOVIRIDAE; ROTAVIRUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
TX MEDLINE; 90080150.  
FA GORTIGLIA M., NISHIKAWA K., HOSHINO Y., TANIGUCHI K.;  
PT "Similarity of the outer capsid protein VP4 of the Gottfried strain  
of porcine rotavirus to that of asymptomatic human rotavirus  
strains.";  
AT J. VIROL. 64:414-418(1990).  
CC -1- SUBCELLULAR LOCATION: OUTER CAPSID.  
CC -1- PTM: VP8 IS ONE OF TWO TRYPSIN CLEAVAGE PRODUCTS OF VP4; THE  
OTHER PRODUCT IS VP5.  
CC -1- SIMILARITY: BELONGS TO THE ROTAVIRUS VP4 PROTEIN FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
DB EMBL; N33516; G333316;  
DR PIR; A33567; VPXSPG.  
DR PFAM; PF00426; VP4; 1.  
KW COAT PROTEIN; GLYCOPROTEIN.  
FT CHAIN 1 775  
FT CHAIN 1 240  
FT CHAIN 247 775  
FT CARBOHYD 32 32  
FT CARBOHYD 56 56  
FT CARBOHYD 97 97  
FT CARBOHYD 132 132  
FT CARBOHYD 324 324  
FT CARBOHYD 583 583  
SQ SEQUENCE 775 AA; 87098 MW; 9F77E993 CRC32;  
Query Match 10.7%; Score 86; DB 1; Length 775;  
Best Local Similarity 36.4%; Pred. No. 5.92e-01;  
Matches 12; Conservative 8; Mismatches 12; Indels 1; Gaps 1;  
Db 215 CNEYINNGLPP-IQONTRNVVPLSLSSRSSTFNK 246  
QY 72 CNGHONPALQOQTHSSRQVTPLSRSSTFNK 104

HUMAN ROTAVIRUS: (SEROTYPE 2 / STRAIN DSI).

VIRUSES; DSRNA VIRUSES; REOVIRIDAE; ROTAVIRUS.

[1]

SEQUENCE FROM N.A.  
MEDLINE: 86275070.  
GORIGLIA M., GREEN K., NISHIKAWA K., TANIGUCHI K., JONES R.,  
KAPIKAN A.Z., CHANOCK R.M.;  
"Sequence of the fourth gene of human rotaviruses recovered from  
asymptomatic or symptomatic infections.";  
J. VIROL. 62:2978-2984(1988).  
[2]  
SEQUENCE OF 1-280.  
MEDLINE: 86313706.  
GORIGLIA M., HOSHINO Y., BUCKLER-WHITE A., BLUMENTALS I., GLASS R.,  
FLORES J.J., KAPIKAN A.Z., CHANOCK R.M.;  
"Conservation of amino acid sequence of VP8 and cleavage region of  
84-kDa outer capsid protein among rotaviruses recovered from  
asymptomatic neonatal infection.";  
PROC. NATL. ACAD. SCI. U.S.A. 83:7039-7043(1986).  
CC -!- SUBCELLULAR LOCATION: OUTER CAPSID.  
CC -!- PTM: VP8 IS ONE OF TWO TRYPSIN CLEAVAGE PRODUCTS OF VP4; THE  
OTHER PRODUCT IS VP5.  
CC -!- SIMILARITY: BELONGS TO THE ROTAVIRUS VP4 PROTEIN FAMILY.  
DR PIR: B28839; VPKRW4.  
DR PIR: C25904; VPKRDS.  
KW PFAM: PF00426; VP4; 1.  
COAT PROTEIN; GLYCOPROTEIN.  
FT CHAIN 1 775  
OUTER CAPSID PROTEIN VP4.  
FT CHAIN 1 240  
OUTER CAPSID PROTEIN VP8.  
FT CARBOHYD 56 56  
POTENTIAL.  
FT CARBOHYD 132 132  
POTENTIAL.  
FT CARBOHYD 150 150  
POTENTIAL.  
FT CARBOHYD 195 195  
POTENTIAL.  
FT CARBOHYD 324 324  
POTENTIAL.  
FT CARBOHYD 583 583  
POTENTIAL.  
FT CARBOHYD 589 589  
POTENTIAL.  
FT CARBOHYD 592 592  
POTENTIAL.  
FT CARBOHYD 599 599  
POTENTIAL.  
FT CONFLICT 52 53  
HG -> SW (IN REF. 2).  
FT CONFLICT 106 107  
IA -> SS (IN REF. 2).  
FT CONFLICT 142 144  
MKF -> TET (IN REF. 2).  
FT CONFLICT 245 245  
R -> K (IN REF. 2).  
FT CONFLICT 280 280  
V -> I (IN REF. 2).  
SQ SEQUENCE 775 AA; 87592 MW; C01DD686 CRC32;  
  
Query Match 10.7%; Score 86; DB 1; Length 775;  
Best Local Similarity 39.4%; Pred. No. 5.92e-01;  
Matches 13; Conservative 6; Mismatches 13; Indels 1; Gaps

DB 215 CNEYINNGLPP-IQNTNRNVPLSLSSRSIOYRR 246  
II : I ::::: IIII II :  
OY 72 CNGHNPALQQGTHSSRQVTPLSRSSSTFNK 104

RESULT 9  
IC ID XJVL YEAST STANDARD; PRT; 147 AA.  
AC AD Q0898;  
DT DT 01-FEB-1995 (REL. 31, CREATED)  
DT DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)  
DT DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE DE VERY HYPOTHEICAL 16.8 KD PROTEIN IN CRT1 REGION.  
GN GN YJL211C OR J0238 OR HD147.  
OS OS SACCHAROMYCES CREVISIAE (BAKER'S YEAST).  
OC OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMICETES; SACCHAROMYCETALES;  
OC OC SACCHAROMYCETACEAE; SACCHAROMYCES.  
RN RN [1];  
RN RN SEQUENCE FROM N.A.  
RC RC STRAIN-S288C;  
RX RX MEDLINE: 95242842.  
RX VAN DENBOL M., DURAND P., BOLLE P.-A., DION C., PORTETELLE D.,  
HILGER F.;  
RA "Sequene analysis of a 4.0 kb DNA fragment located near the left

CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

Q5 LACTOBACILLUS DELBRUECKII (SUBSP. BULGARICUS).  
Q6 BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; LACTOBACILLACEAE;  
Q7 LACTOBACILLUS.

STRAIN=A1CC 1184Z;  
MEDLINE: 94304994.

TRANSAMINASE BY INDUCTION. ;  
RL BIOCHIMIE 76:45-49(1994).

**CC** THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
**CC** between the Swiss Institute of Bioinformatics and the EMBL Outstation  
**CC** at the European Bioinformatics Institute. There are no restrictions on its  
**CC** use by non-profit institutions as long as its content is, in no way  
**CC** modified and this statement is not removed. Usage by and for commercial  
**CC** entities requires a license agreement (See [http://www.isb-sib.ch/announcement/  
or\\_send\\_an\\_email\\_to\\_license@isb-sib.ch](http://www.isb-sib.ch/announcement/or_send_an_email_to_license@isb-sib.ch)).  
**CC**

PMRI: V61190: C43095

DR	PROSITE; PS00595; AA_TRANSFER_CLASS_5; 1.
DR	PFAM; PF00266; aminotran_5; 1.
KV	PYRIDOXAL PHOSPHATE. PYRIDOXAL PHOSPHATE. 191 191 FT BINDING 355 355 FT NON_TER 355 355 EQ SEQUENCE 355 AA; 38804 MW; EE00BD60 CRC32; PYPIDOXAL PHOSPHATE (BY SIMILARITY).

234 AMAKAR 240  
:::  
92 PLSR 98

01-NOV-1990 (REL. 16, CREATED)  
01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)  
01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)  
DE PHOSPHATIDYLCHOLINE-STEROL ACYLTRANSFERASE  
DE (LECITHIN-CHOLESTEROL ACYLTRANSFERASE) (PHOSPHOLIPID-CHOLESTEROL  
DE ACYLTRANSFERASE)  
GN I.CAT.

OS RATTUS NORVEGICUS (RAT).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.

SEQUENCE FROM N. A.  
MEDLINE: 90384859.



```

DR PIR: D28839; VPXRW6.
DR PFAM: PF00426; VP4; 1.
KW COAT PROTEIN; GLYCOPROTEIN.
FT CHAIN 1 775
FT CHAIN 1 240
FT CHAIN 247 775
FT CARBOHYD 32 56
FT CARBOHYD 56 56
FT CARBOHYD 97 97
FT CARBOHYD 129 129
FT CARBOHYD 132 132
FT CARBOHYD 132 132
FT CARBOHYD 195 195
FT CARBOHYD 324 324
FT CARBOHYD 577 577
FT CARBOHYD 583 583
FT CARBOHYD 589 589
FT CARBOHYD 592 592
SQ SEQUENCE 775 AA; 87609 MW; 23C46EA5 CRC32;

Query Match 10.5%; Score 85; DB 1; Length 775;
Best Local Similarity 36.4%; Pred. No. 8.32e-01;
Matches 12; Conservative 8; Mismatches 12; Indels 1; Gaps 1;

Db 215 CNEYINNGLPP-IQNRNVVPLSLRSIQYR 246
QY 72 CNGHQPALQOQTHSRQVTPLSLRSSTFNK 104

RESULT 15
ID VP4_ROTH1 STANDARD; PRT: 775 AA.
AC P11198;
DT 01-JUL-1989 (REL. 11, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
DE OUTER CAPSID PROTEIN VP4 (HEMAGGLUTININ) (OUTER LAYER PROTEIN VP4)
DE [CONTAINS: OUTER CAPSID PROTEINS VP5 AND VP8].
GN S4.
OS HUMAN ROTAVIRUS (SEROTYPE 1 / STRAIN 1076).
OC VIRUSES; DSRNA VIRUSES; REOVIRIDAE; ROTAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 88275070.
RA GORZIGLIA M., GREEN K., NISHIKAWA K., TANIGUCHI K., JONES R.,
RA KAPIKIAN A.Z., CHANOCK R.M.;
RT "Sequence of the fourth gene of human rotaviruses recovered from
RT asymptomatic or symptomatic infections.";
RL J. VIROL. 62:2978-2984(1988).
RN [2]
RP SEQUENCE OF 1-280.
RX MEDLINE; 86313706.
RA GORZIGLIA M., HOSHINO Y., BUCKLER-WHITE A., BLOUMENTALS I., GLASS R.,
RA FLORES J., KAPIKIAN A.Z., CHANOCK R.M.;
RT "Conservation of amino acid sequence of VP8 and cleavage region of
RT 84-kda outer capsid protein among rotaviruses recovered from
RT asymptomatic neonatal infection.";
RL PROC. NATL. ACAD. SCI. U.S.A. 83:7039-7043(1986).
CC -!- SUBCELLULAR LOCATION: OUTER CAPSID.
CC -!- PTM: VP8 IS ONE OF TWO TRYPSIN CLEAVAGE PRODUCTS OF VP4; THE
CC OTHER PRODUCT IS VP5.
CC -!- SIMILARITY: BELONGS TO THE ROTAVIRUS VP4 PROTEIN FAMILY.
DR PIR: F28839; VPXRW8.
DR PIR: F25904; VPXR16.
DR PFAM: PF00426; VP4; 1.
KW COAT PROTEIN; GLYCOPROTEIN.
FT CHAIN 1 775
FT CHAIN 1 240
FT CHAIN 247 775
FT CARBOHYD 32 32
FT CARBOHYD 56 56
FT CARBOHYD 97 97
FT CARBOHYD 111 111
FT CARBOHYD 114 114
FT CARBOHYD 132 132

```

```

FT CARBOHYD 192 192
FT CARBOHYD 277 277
FT CARBOHYD 324 324
FT CARBOHYD 583 583
FT CARBOHYD 606 606
SQ SEQUENCE 775 AA; 87697 MW; 55440B49 CRC32;

Query Match 10.5%; Score 85; DB 1; Length 775;
Best Local Similarity 33.3%; Pred. No. 8.32e-01;
Matches 11; Conservative 9; Mismatches 12; Indels 1; Gaps 1;

Db 215 CNEYINNGLPP-MQNRNIVPVALSSRSVYQR 246
QY 72 CNGHQPALQOQTHSRQVTPLSLRSSTFNK 104

Search completed: Wed Dec 22 10:23:29 1999
Job time : 9 secs.

```



\*\*\*\*\*  
WATERMAN  
\*\*\*\*\*  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Wed Dec 22 10:23:46 1999; MasPar time 13.00 Seconds  
Tabular output not generated. 436.538 Million cell updates/sec

Title: >US-09-209-961-9  
Description: (1-104) from US09209961.pep  
Perfect Score: 807  
Sequence: 1 MVTIPPLVPRWPCGFRVC.....HSSRQVTPLSLRSSTFNK 104

Scoring table: PAM 150  
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: sptrembl9  
1:sp.archaea 2:sp.bacteria 3:sp.fungi 4:sp.human  
5:sp.invertebrate 6:sp.mammal 7:sp.mhc 8:sp.organelle  
9:sp.phage 10:sp.plant 11:sp.rodent 12:sp.unclassified  
13:sp.vertebrate 14:sp.virus

Statistics: Mean 39.165; Variance 73.303; scale 0.534

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					
Result	Query	Match	Length	ID	Description
No.	Score				Pred. No.
1	807	100.0	104	14	O92286 P12.1. 1.69e-148
2	761	94.3	104	14	O56124 STRAIN PMWS PCV, COMPL 2.48e-138
3	735	91.1	104	14	O93186 ORF3. 1.34e-132
4	104	12.9	289	3	O05521 SIMILARITY TO NADH-UBI 6.10e-03
5	100	12.4	292	11	O35839 LECITHIN:CHOLESTEROL A 2.42e-02
6	99	12.3	379	5	O21483 M03C11.1 PROTEIN. 3.39e-02
7	92	11.4	413	3	P78759 FISSION YEAST (FRAGMENT 3.46e-01
8	92	11.4	711	5	O45825 T25E12.11 PROTEIN. 3.46e-01
9	91	11.3	507	8	O20281 MATURASE (FRAGMENT). 4.78e-01
10	90	11.2	82	2	O45911 ORF 109. 6.59e-01
11	90	11.2	103	2	O28873 HYPOTHETICAL 12.1 KD P 6.59e-01
12	90	11.2	499	8	O30533 MATURASE (FRAGMENT). 6.59e-01
13	90	11.2	505	8	O20276 MATURASE (FRAGMENT). 6.59e-01
14	89	11.0	150	14	O56350 GP41 (FRAGMENT). 9.06e-01
15	89	11.0	397	4	P78423 CX3C CHEMOKINE PRECURS 9.06e-01
16	89	11.0	506	8	O78332 MATURASE (FRAGMENT). 9.06e-01
17	89	11.0	507	8	O47164 RIBOSOMAL MATURASE. 9.06e-01
18	89	11.0	511	10	O81939 MATURASE (FRAGMENT). 9.06e-01
19	89	11.0	541	5	O09617 HYPOTHETICAL 61.0 KD P 9.06e-01
20	89	11.0	700	2	O53344 PUTATIVE DNA HELICASE 9.06e-01

21	89	11.0	787	4	O99835 SMOOTHENED. 9.06e-01
22	89	11.0	793	11	O70246 PUTATIVE NEURONAL CELL 9.06e-01
23	89	11.0	793	11	P97698 SMOOTHENED. 9.06e-01
24	88	10.9	273	14	P67537 GROUP A ROTAVIRUS ISOL 1.24e+00
25	82	10.9	455	5	O19442 SIMILARITY WITH VARIOU 1.24e+00
26	88	10.9	775	14	O86185 VP5 AND VP8. 1.24e+00
27	88	10.9	775	14	O82012 VP4 PROTEIN. 1.24e+00
28	88	10.9	775	14	O86111 OUTER CAPSID VP4. 1.24e+00
29	87	10.8	274	5	O21659 R03A10.6 PROTEIN (FRAG 1.70e+00
30	87	10.8	291	11	O35502 LECITHIN:CHOLESTEROL A 1.70e+00
31	87	10.8	298	11	O35726 LECITHIN:CHOLESTEROL A 1.70e+00
32	87	10.8	299	11	O35573 LECITHIN:CHOLESTEROL A 1.70e+00
33	87	10.8	511	10	O81935 MATURASE (FRAGMENT). 1.70e+00
34	87	10.8	631	10	O90844 F4118.27 PROTEIN. 1.70e+00
35	96	10.7	259	5	O45911 Y17D7B.3 PROTEIN. 2.33e+00
36	96	10.7	291	11	O35746 LECITHIN:CHOLESTEROL A 2.33e+00
37	86	10.7	421	2	O56906 RFBH. 2.33e+00
38	86	10.7	515	10	O81938 MATURASE (FRAGMENT). 2.33e+00
39	86	10.7	534	10	O96303 PHOSPHATE TRANSPORTER. 2.33e+00
40	85	10.7	775	14	O82118 RNA FOR VP4. 2.33e+00
41	95	10.5	288	2	P73993 HYPOTHETICAL 33.1 KD P 3.17e+00
42	85	10.5	344	5	O45991 ZK1037.9 PROTEIN. 3.17e+00
43	85	10.5	440	11	O35849 LECITHIN:CHOLESTEROL A 3.17e+00
44	35	10.5	507	8	O62949 MATURASE (FRAGMENT). 3.17e+00
45	95	10.5	803	5	O22343 T08D10.2 PROTEIN. 3.17e+00

ALIGNMENTS

RESULT 1  
ID O92286 PRELIMINARY; PRT; 104 AA.  
AC O92286;  
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)  
TT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE P12.1.  
OS PORCINE CIRCOVIRUS.  
OC VIRUSES; SSDNA VIRUSES; CIRCOVIRIDAE; CIRCOVIRUS.  
KW [1]  
RF SEQUENCE FROM N.A.  
RC STRAIN=412;  
RA WANG L., WILLSON P., CHOW B., GIBBONS E., BABIUK L.;  
RT "Emergence of a new porcine circovirus."  
RL SUBMITTED (AUG-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: AF085695; G3668368;  
SQ SEQUENCE 10, AA; 12060 MW; 820A1C5D CRC32;

Query Match 100.0%; Score 807; DB 14; Length 104;  
Best Local Similarity 100.0%; Pred. No. 1.69e-148;  
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 MVTIPPLVPRWPCGFRVCISPPAFPTTPRPVPHNEVIGFPTILLHFFPAHFQKFSQPA 60  
QV 1 MVTIPPLVPRWPCGFRVCISPPAFPTTPRPVPHNEVIGFPTILLHFFPAHFQKFSQPA 60  
Db 61 EITDKRIRVLLCNGHONPALQOQTHSSRQVTPLSLRSRSTFNK 104  
QV 61 EITDKRIRVLLCNGHONPALQOQTHSSRQVTPLSLRSRSTFNK 104

RESULT 2  
ID O56124 PRELIMINARY; PRT; 104 AA.  
AC O56124;  
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
TT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE STRAIN PMWS PCV, COMPLETE GENOME.  
OS PORCINE CIRCOVIRUS.  
OC VIRUSES; SSDNA VIRUSES; CIRCOVIRIDAE; CIRCOVIRUS.  
KW [1]  
RF SEQUENCE FROM N.A.  
RC STRAIN=PMWS PCV;  
DE MEDLINE; 98241772.

RA HAMEL A.L., LIN L.L., NAYAR G.P.S.;  
RT "Nucleotide sequence of porcine circovirus associated with  
RT postweaning multisystemic wasting syndrome in pigs.";  
RL J. VIROL. 72:5262-5267(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-PWIS PCV;  
RA HAMEL A.L., LIN L.L., NAYAR G.P.S.;  
RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-PORCINE CIRCOVIRUS TYPE II;  
RX MEDLINE; 98418498  
RA MEHRAN B.M., MCNEILLY F.M., TODD D., KENNEDY S., JEWHRST V.,  
RA ELLIS J.A., HASSARD L.E., CLARK E.G., HAINES D.M., ALLAN G.M.;  
RT "Characterisation of novel circovirus DNAs associated with wasting  
RT syndromes in pigs.";  
RL J. GEN. VIROL. 79:2171-2199(1998).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN-PORCINE CIRCOVIRUS TYPE II;  
RA MEHRAN B.M.;  
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; AF027217; G2689648;  
DR EMBL; AF055392; G3598811;  
DR EMBL; AF055393; G3598799;  
SQ SEQUENCE 104 AA; 11879 MW; 68446EFF CRC32;  
  
Query Match 94.3%; Score 761; DB 14; Length 104;  
Best Local Similarity 95.2%; Pred. No. 2.48e-138;  
Matches 99; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
  
Db 1 MTIPLVRFWFPVCGFRVCKISSPFAFTTPRPHNDVYIGLPTLLHFFAFOKFSQPA 60  
QY 1 MTIPLVRFWFPVCGFRVCKISSPFAFTTPRPHNDVYIGLPTLLHFFAFOKFSQPA 60  
  
Db 61 EISDKRYVLLCNHQTALQOQTHSSRQVTPLSLRSRSTFNK 104  
QY 61 EIPDKRYVLLCNHQTALQOQTHSSRQVTPLSLRSRSTFNK 104  
  
RESULT 3  
ID O93186 PRELIMINARY; PRT; 104 AA.  
AC O93186;  
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE ORF3  
OS PORCINE CIRCOVIRUS.  
OC VIRUSES; SS DNA VIRUSES; CIRCOVIRIDAE; CIRCOVIRUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-PORCINE CIRCOVIRUS TYPE II;  
RX MEDLINE; 98418498  
RA MEHRAN B.M., MCNEILLY F.M., TODD D., KENNEDY S., JEWHRST V.,  
RA ELLIS J.A., HASSARD L.E., CLARK E.G., HAINES D.M., ALLAN G.M.;  
RT "Characterisation of novel circovirus DNAs associated with wasting  
RT syndromes in pigs.";  
RL J. GEN. VIROL. 79:2171-2199(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-PORCINE CIRCOVIRUS TYPE II;  
RA MEHRAN B.M.;  
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; AF055394; G3598834;  
DR EMBL; AF055393; G3598823;  
SQ SEQUENCE 104 AA; 11898 MW; 01C3FD3F CRC32;  
  
Query Match 91.1%; Score 735; DB 14; Length 104;  
Best Local Similarity 92.2%; Pred. No. 1.34e-132;  
Matches 95; Conservative 5; Mismatches 3; Indels 0; Gaps 0;  
  
Db 1 MTIPLVRFWFPVCGFRVCKISSPFAFTTPRPHNDVYISLPTLLHFFAFOKFSQPA 60

QY 1 MTIPLVRFWFPVCGFRVCKISSPFAFTTPRPHNDVYIGLPTLLHFFAFOKFSQPA 60  
Db 61 EISDKRYVLLCNHQTALQOQTHSSRQVTPLSLRSRSTLNH 103  
QY 61 EIPDKRYVLLCNHQTALQOQTHSSRQVTPLSLRSRSTFN 103  
  
RESULT 4  
ID O05521 PRELIMINARY; PRT; 289 AA.  
AC O05521;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
DE SIMILARITY TO NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 4.  
GN D9819.10.  
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOCYCETES; SACCHAROMYCETALES;  
OC SACCHAROMYCETACEAE; SACCHAROMYCETES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-S288C (AB972);  
RA FULTON L.;  
RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-S288C (AB972);  
RA JOHNSTON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DU Z.,  
RA FAVELLO A., FULTON L., GATTUNG S., GRECO T., KIRSTEN J., KUCABA T.,  
RA HALLSWORTH K., HAWKINS J., HILLIER L., JIER M., JOHNSON D.,  
RA JOHNSTON L., LANGSTON Y., LATREILLE P., LE T., MARDIS E., MENEZES S.,  
RA MILLER N., NHAN M., PAULEY A., PELUSO D., RIFKEN L., RILES L.,  
RA TAICH A., TREVASKIS E., VIGNATI D., WILCOX L., WOHLDMAN P., VAUDIN M.,  
RA WILSON R., WATERSTON R.;  
RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-S288C (AB972);  
RA JOHNSTON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DU Z.,  
RA FAVELLO A., FULTON L., GATTUNG S., GRECO T., KIRSTEN J., KUCABA T.,  
RA HALLSWORTH K., HAWKINS J., HILLIER L., JIER M., JOHNSON D.,  
RA JOHNSTON L., LANGSTON Y., LATREILLE P., LE T., MARDIS E., MENEZES S.,  
RA MILLER N., NHAN M., PAULEY A., PELUSO D., RIFKEN L., RILES L.,  
RA TAICH A., TREVASKIS E., VIGNATI D., WILCOX L., WOHLDMAN P., VAUDIN M.,  
RA WILSON R., WATERSTON R.;  
RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN-S288C (AB972);  
RA WATERSTON R.;  
RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [5]  
RP SEQUENCE FROM N.A.  
RC STRAIN-S288C (AB972);  
RA JIA Y., CHERRY J.M.;  
RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; U51031; G1332640;  
KW UBIQUINONE.  
SQ SEQUENCE 289 AA; 33514 MW; 1A871897 CRC32;  
  
Query Match 12.9%; Score 104; DB 3; Length 289;  
Best Local Similarity 29.7%; Pred. No. 6.10e-03;  
Matches 11; Conservative 13; Mismatches 12; Indels 1; Gaps 1;  
  
Db 3 RVSEIKTPFNIGA-KWLEDFVLLIIMLLNYPVYQ 38  
QY 18 RVCKISSPFAFTTPRPHNDVYIGLPTLLHFFAFOKFSQPA 54  
  
RESULT 5  
ID O35839 PRELIMINARY; PRT; 292 AA.  
AC O35839;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
DE LECITHIN:CHOLESTEROL ACYL TRANSFERASE (FRAGMENT).  
GN LCAT.  
OS SCIURUS GRISEUS.  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
OC SCIURIGNATHI; SCIURIDAE; SCIURINAE; SCIURUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA ROBINSON M., CATZEFIS F.M., BRIOLAY J., MOUCHIROUD D.,  
RL SUBMITTED (SEP-1996) TO EMBL/GENBANK/DBJ DATA BANKS.



OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;  
 OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;  
 OC ASTERIDAE; GENTIANAE; GENTIANACEAE; APOCYNACEAE; STROPHANTHUS.  
 RN [1]

RA SEQUENCE FROM N.A.  
 RA ENDRESS M.E., SENBLAD B., NILSSON S., CIVEYREL L., CHASE M.W.,  
 RA HUYSMANS S., GRAFSTROEM E., BREMER B.;  
 RL OPERA BOT. BELG. 7:59-102(1996).  
 DR EMBL: Z70186; E332376; -.  
 KW CHLOROPLAST.  
 FT NON\_TER 1  
 SQ SEQUENCE 507 AA; 59810 MW; 6B93832E CRC32;

Query Match 11.3%; Score 91; DB 8; Length 507;  
 Best Local Similarity 29.5%; Pred. No. 4.78e-01;  
 Matches 18; Conservative 11; Mismatches 30; Indels 2; Gaps 2;

Db 209 FLLNFYVCEYESIFVFLHNSHLRP-TSGV-LLEIRYFYGKIERLAETVDFQANL 266  
 Qy 12 FVCGFRVCKISSPFAFTTPRPHNEVIGPITLLHPFAHFQKFSQPAEIDFKRYVLL 71

Db 267 C 267  
 Qy 72 C 72

RESULT 10  
 ID Q45911 PRELIMINARY; PRT; 82 AA.

AC Q45911;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)  
 DE ORF 109.  
 OS COXIELLA BURNETII.  
 OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; COXIELLA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-RUSSIAN ISOLATE Q 1182;  
 RA WILLEMS H., THIELE D., VALKOVA D.;  
 RL SUBMITTED (FEB-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: X84722; G682755; -.  
 SQ SEQUENCE 82 AA; 9545 MW; FE63A40C CRC32;

Query Match 11.2%; Score 90; DB 2; Length 82;  
 Best Local Similarity 20.8%; Pred. No. 6.59e-01;  
 Matches 10; Conservative 18; Mismatches 19; Indels 1; Gaps 1;

Db 28 LLLVPSLVLFWMFPISSIDAYLIVQYGLAIWLCVLDLYTYLLASL 75  
 Qy 1 MVTIPPLV-FRWFPVCGFRVCKISSPFAFTTPRPHNEVIGPITLL 47

RESULT 11  
 ID Q52873 PRELIMINARY; PRT; 103 AA.

AC Q52873;  
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
 DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
 DE HYPOTHETICAL 12.1 KD PROTEIN.  
 OS COXIELLA BURNETII.  
 OC PLASMID OPRS.  
 OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; COXIELLA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-PRISCILLA Q177;  
 RA LAUTENSCHLAGER S., JAEGER C., WILLEMS H., BALJER G.;  
 RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: Y15898; E1216848; -.  
 KW HYPOTHETICAL PROTEIN; PLASMID.  
 SQ SEQUENCE 103 AA; 12085 MW; 52F4B0F9 CRC32;

Query Match 11.2%; Score 90; DB 2; Length 103;  
 Best Local Similarity 20.8%; Pred. No. 6.59e-01;

Matches 10; Conservative 18; Mismatches 19; Indels 1; Gaps 1;  
 Db 28 LLLVPSLVLFWMFPISSIDAYLIVQYGLAIWLCVLDLYTYLLASL 75  
 Qy 1 MVTIPPLV-FRWFPVCGFRVCKISSPFAFTTPRPHNEVIGPITLL 47

RESULT 12  
 ID O03053 PRELIMINARY; PRT; 499 AA.

AC O03053;  
 DT 01-JUL-1997 (TREMBLREL. 04, CREATED)  
 DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)  
 DT 01-JUL-1997 (TREMBLREL. 04, LAST ANNOTATION UPDATE)  
 DE MATURASE (FRAGMENT).  
 GN MATK.  
 OS PICRALIMA NITIDA.  
 OC CHLOROPLAST.  
 OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;  
 OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;  
 OC ASTERIDAE; GENTIANAE; GENTIANACEAE; APOCYNACEAE; PICRALIMA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA ENDRESS M.E., SENBLAD B., NILSSON S., CIVEYREL L., CHASE M.W.,  
 RA HUYSMANS S., GRAFSTROEM E., BREMER B.;  
 RL OPERA BOT. BELG. 7:59-102(1996).  
 DR EMBL: Z70179; E228637; -.  
 KW CHLOROPLAST.  
 FT NON\_TER 1  
 SQ SEQUENCE 499 AA; 58904 MW; 53AB1C71 CRC32;

Query Match 11.2%; Score 90; DB 8; Length 499;  
 Best Local Similarity 29.5%; Pred. No. 6.59e-01;  
 Matches 18; Conservative 11; Mismatches 30; Indels 2; Gaps 2;

Db 203 FFLNFYVCEYESIFVFLHNSHLRS-TSGV-LLEIRYFYGKIERLVEVEAKDFOVNL 260  
 Qy 12 FVCGFRVCKISSPFAFTTPRPHNEVIGPITLLHPFAHFQKFSQPAEIDFKRYVLL 71

Db 261 C 261  
 Qy 72 C 72

RESULT 13  
 ID O20276 PRELIMINARY; PRT; 505 AA.

AC O20276;  
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE MATURASE (FRAGMENT).  
 GN MATK.  
 OS PICRALIMA NITIDA.  
 OC CHLOROPLAST.  
 OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;  
 OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;  
 OC ASTERIDAE; GENTIANAE; GENTIANACEAE; APOCYNACEAE; PICRALIMA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA ENDRESS M.E., SENBLAD B., NILSSON S., CIVEYREL L., CHASE M.W.,  
 RA HUYSMANS S., GRAFSTROEM E., BREMER B.;  
 RL OPERA BOT. BELG. 7:59-102(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA CIVEYREL L.;  
 RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA CIVEYREL L.;  
 RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: Z70179; E329795; -.  
 KW CHLOROPLAST.  
 FT NON\_TER 1  
 SQ SEQUENCE 505 AA; 59552 MW; C87B6754 CRC32;

Query Match 11.2%; Score 90; DB 8; Length 505;  
 Best Local Similarity 29.5%; Pred. No. 6.59e-01;  
 Matches 18; Conservative 11; Mismatches 30; Indels 2; Gaps 2;  
 Db 209 FFLNFFVCEVESIFVLHNOSSHLRS-TSGV-LLERIYGYKIERLVEVFAKDFQVNL 266  
 QY 12 FVCGFRVCKISSPFAFTTPRMPHNEVIGPITLLHPAHPFQKFSQPAEIDKRYRVL 71  
 Db 267 C 267  
 QY 72 C 72

RESULT 14  
 ID O56350 PRELIMINARY; PRT; 150 AA.  
 AC O56350;  
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
 DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)  
 DE GP41 (FRAGMENT).  
 GN ENV.  
 OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 2 (HIV-2).  
 OC VIRUSES; RETROVIRUSES; RETROVIRIDAE; LENTIVIRUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-PO2;  
 RA HEREDIA A.;  
 RL AIDS RES. HUM. RETROVIRUSES 0:0-0(1998).  
 DR EMBL; AF039491; G2828072;  
 FT NON\_TER 1  
 FT NON\_TER 150 150  
 SQ SEQUENCE 150 AA; 17976 MW; 23720A57 CRC32;

Query Match 11.0%; Score 89; DB 14; Length 150;  
 Best Local Similarity 47.6%; Pred. No. 9.06e-01;  
 Matches 10; Conservative 5; Mismatches 4; Indels 2; Gaps 2;  
 Db 42 CAFROVCHTTPWFNSTPKW 62  
 QY 15 CGFR-VCKISSPFAF-TTPRW 33

RESULT 15  
 ID P78423 PRELIMINARY; PRT; 397 AA.  
 AC P78423; O00672;  
 DT 01-MAY-1997 (TREMBLREL. 03, CREATED)  
 DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE CX3C CHEMOKINE PRECURSOR.  
 GN A-152E5.2.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
 OC CATARRHINI; HOMINIDAE; HOMO.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 9717111.  
 RA BAZAN J.F., BACON K.B., HARDIMAN G., WANG W., SOO K., ROSSI D.,  
 RA GREAVES D.R., ZLOTNIK A., SCHALL T.J.;  
 RT "A new class of membrane-bound chemokine with a CX3C motif.";  
 RL NATURE 385:640-644(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA ADAMS M.D., LOFTUS B.J., ZHOU L., CROSBY M., FUHRMANN J., MASON T.M.,  
 RA BRANDON R., KIM U.J., KERLAVAGE A.R., VENTER J.C.;  
 RT "Homo sapiens Chromosome 16 BAC clone CIT987SK-A-152E5.";  
 RL SUBMITTED (JUN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL; U91835; G1895259;  
 DR EMBL; U84487; G1888523;  
 DR EMBL; AC004382; G3252821;  
 DR PFAM; PF00048; 118; 1.  
 KW SIGNAL.  
 FT SIGNAL 1 24 POTENTIAL.

FT CHAIN 25 397 CX3C.CHEMOKINE.  
 SQ SEQUENCE 397 AA; 42202 MW; C8093D7D CRC32;  
 Query Match 11.0%; Score 89; DB 4; Length 397;  
 Best Local Similarity 26.2%; Pred. No. 9.06e-01;  
 Matches 17; Conservative 21; Mismatches 23; Indels 4; Gaps 4;  
 Db 43 IPVALLIHYQNOASCGRKRAIILETRHRLFCADPKQWVKDAMQHLDRAAALT-RN-G 100  
 QY 42 FPITLL-HFPAHFQKFSQPAEIDKRYRVLNCGHNPALQOQT-HSSRQVTPLSLRSRS 99  
 Db 101 GTFEK 105  
 QY 100 STFNK 104

Search completed: Wed Dec 22 10:24:02 1999  
 Job time : 16 secs.

\*\*\*\*\*

W P E R L H

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Dec 22 10:11:23 1999; MasPar time 20.50 Seconds  
Tabular output not generated. 325.700 Million cell updates/sec

Title: >US-09-209-961-3  
Description: (1-314) from US09209961.pep  
Perfect Score: 2358  
Sequence: 1 MPSKNGRSGPQPKRWFT.....GGQFVTLSPCPPEPYEINY 314  
Scoring table: PAM 150  
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq35

Statistics: Mean 34.382; Variance 135.259; scale 0.254

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	197	8.4	286	38 W87459	BBTV DNA I clone (7-4	2.21e-08
2	193	8.2	285	38 W87460	BBTV DNA II clone (2-	5.24e-08
3	139	5.9	286	38 W87461	BBTV DNA II V2 Cl ami	4.05e-03
4	136	5.8	286	31 W54071	Banana bunchy top vir	7.38e-03
5	97	4.1	611	19 W02157	Periplasmic Beta-N-ac	1.22e+01
6	97	4.1	611	39 W85599	Hexosaminidase enzyme	1.22e+01
7	92	3.9	165	34 W62673	Streptococcus pneumoniae	2.94e+01
8	91	3.9	820	9 R47468	Branching enzyme of r	3.50e+01
9	91	3.9	820	11 R53228	Rice starch branching	3.50e+01
10	92	3.9	1822	10 R52273	Beta subunit of integ	2.94e+01
11	93	3.9	2150	3 P60045	Sequence of viral pro	2.47e+01
12	93	3.9	2150	2 P81048	Sequence encoded by R	2.47e+01
13	93	3.9	2150	2 R05127	Complete human rhinov	2.47e+01
14	92	3.9	2227	3 P60066	Sequence of viral L43	2.94e+01
15	92	3.9	2227	27 W34074	Hepatitis A virus HM-	2.94e+01
16	89	3.8	357	9 R44618	Human retinol binding	4.95e+01

17	89	3.8	445	27 W27603	Rat neuropeptide Y re	4.95e+01
18	89	3.8	445	22 W15232	Rat neuropeptide Y-Y5	4.95e+01
19	85	3.8	456	25 W29446	Rat hypothalamic neur	4.95e+01
20	85	3.8	456	29 W37092	Rattus norvegicus hyp	4.95e+01
21	89	3.8	456	18 R95940	Rat Y5 receptor.	4.95e+01
22	89	3.8	456	26 W29412	Rat hypothalamic neur	4.95e+01
23	89	3.8	466	22 W15233	Mouse neuropeptide Y-	4.95e+01
24	89	3.8	507	3 P50263	Sequence encoded by P	4.95e+01
25	90	3.8	830	37 W80314	CIITA (class II trans	4.17e+01
26	90	3.8	987	31 W57057	Class II trans activa	4.17e+01
27	90	3.8	1106	37 W80312	CIITA (class II trans	4.17e+01
28	90	3.8	1130	31 W57056	Class II trans activa	4.17e+01
29	90	3.8	1130	14 R74452	Class II transactivat	4.17e+01
30	90	3.8	1130	16 R81569	Class II transactivat	4.17e+01
31	90	3.8	1130	37 W80313	CIITA (class II trans	4.17e+01
32	90	3.8	1207	37 W80311	CIITA (class II trans	4.17e+01
33	90	3.8	1463	6 R29940	Deduced from Lelystad	4.17e+01
34	90	3.8	2227	1 R05697	Attenuated hepatitis	4.17e+01
35	90	3.8	4536	28 W41262	Apolipoprotein B-100.	4.17e+01
36	88	3.7	445	22 W15230	Human neuropeptide Y-	5.88e+01
37	88	3.7	445	27 W27604	Human neuropeptide Y-	5.88e+01
38	88	3.7	455	18 R95939	Human Y5 receptor.	5.88e+01
39	88	3.7	455	29 W37093	Homo sapiens hippocam	5.88e+01
40	88	3.7	455	25 W29447	Human hippocampal neu	5.88e+01
41	88	3.7	455	26 W29413	Human hippocampal neu	5.88e+01
42	88	3.7	533	9 R44617	Human retinol binding	5.88e+01
43	86	3.6	398	31 W37988	Amino acid sequence O	8.27e+01
44	86	3.6	429	26 W40801	Alpha-galactosidase A	8.27e+01
45	86	3.6	429	12 R63234	Human alpha-galactosi	8.27e+01

## ALIGNMENTS

## RESULT 1

ID W87459 standard; Protein; 286 AA.  
AC W87459;  
DT 10-FEB-1999 (first entry)  
DE BBTV DNA I clone (7-4-2) ORF V2 product DNA I-VI.  
KW Banana bunchy top virus; BBTV; PCR technique; plant; tissue; banana;  
SW Musa; phloem damage; aphid; nucleic acid detection.  
CS Banana bunchy top virus.  
FH Key Location/Qualifiers  
FT Misc\_difference 92 /note= "eroded by GAA"  
FT Misc\_difference 153 /note= "eroded by AGA"  
FT US5846705 A.  
PD 08-DEC-1998.  
PP 06-APR-1995: 418071.  
PR 06-APR-1995: US-418071.  
PA (BIOT-) DEV CENT BIOTECHNOLOGY.  
/L Soong S, Wu R, Yeu L;  
DR WPI; 59-059037/05.  
DR N-PSDB: V71832.  
PT Nucleic acids having banana bunchy top virus component sequences -  
PT used to design primers for use in polymerase chain reaction  
PT detection of the virus  
PS Claim 1; Columns 23-26; 27pp; English.  
CC This represents the ORF V2 product (DNA I-VI) of a banana bunchy top  
CC virus (BBTV) DNA I (clone 7-4-2) circular single stranded (css) DNA. The  
CC invention provides nucleic acid sequences associated with BBTV that can  
CC be used in a PCR technique for detecting BBTV. The nucleic acid sequences  
CC (V71832) to V71833 are used as the basis for the construction of PCR  
CC primers, to detect BBTV infection. The PCR technique is used for  
CC detecting BBTV in plant tissues (preferably banana, especially Musa  
CC species). The virus, one of the most important banana species viruses,  
CC causes phloem damage and is transmitted by aphids. PCR detection gives  
CC accurate, reliable and specific determination of absence or presence of  
CC the virus. 286 AA;  
SQ Sequence 286 AA;

Query Match 8.4%; Score 197; DB 38; Length 286;  
Query Local Similarity 36.5%; Pred. No. 2.21e-08;

Matches 42; Conservative 21; Mismatches 47; Indels 5; Gaps 5;

Db 4 pskwcftlnyssaarenfslilkeedvhyavvgdevapatgkhlglgylslkrlrlg 63  
 QY 13 PHRWFTLNPNSEDERKKIRELPISL-FDYFVIGEGNEEGRTPHLQGFANFVKQTEN 71

Db 64 glkkky-gsrahweiaargdeenskysck-qtllilelfgfvngsnkrkisenva 116  
 QY 72 KRW-W-YLGARCHIEKAKGTDQONKEYCSKREGNLLIECGAPRSOG-QRSDUSTAVS 124

RESULT 2  
 ID W87460 standard; Protein; 285 AA.  
 AC W87460;  
 DT 10-FEB-1999 (first entry)  
 DE BBTV DNA II clone (2-17) ORF V2 product DNA II V2(D2).  
 KW Banana bunchy top virus; BBTV; PCR technique; plant; tissue; banana;  
 MU Musa; phloem damage; aphid; nucleic acid detection.  
 OS Banana bunchy top virus.  
 FH Key Location/Qualifiers  
 FT Misc\_difference 49 /note- "encoded by CAG"

PN US5846705-A.  
 PD 08-DEC-1998.  
 PF 06-APR-1995; 418071.  
 PR 06-APR-1995; US-418071.  
 PA (BIOT-) DEV CENT BIOTECHNOLOGY.  
 PI Soong T, Wu R, You L;  
 DR WPI; 99-059037/05.  
 DR N-PSDB; V71833.

PT Nucleic acids having banana bunchy top virus component sequences -  
 PT used to design primers for use in polymerase chain reaction  
 PT detection of the virus  
 PS Claim 12; Columns 25-28; 27pp; English.  
 CC This represents the ORF V2 product [DNA II V2(D2)] of a banana bunchy top  
 CC virus (BBTV) DNA II (clone 2-17) circular single stranded (css) DNA. The  
 CC invention provides nucleic acid sequences associated with BBTV that can  
 CC be used in a PCR technique for detecting BBTV. The nucleic acid sequences  
 CC (V71830 to V71833) are used as the basis for the construction of PCR  
 CC primers, to detect BBTV infection. The PCR technique is used for  
 CC detecting BBTV in plant tissues (preferably banana, especially Musa  
 CC species). The virus, one of the most important banana species viruses,  
 CC causes phloem damage and is transmitted by aphids. PCR detection gives  
 CC accurate, reliable and specific determination of absence or presence of  
 CC the virus.  
 SQ Sequence 285 AA;

Query Match 8.28; Score 193; DB 38; Length 285;  
 Best Local Similarity 37.99; Pred. No. 5.24e-06;  
 Matches 39; Conservative 18; Mismatches 42; Indels 4; Gaps 4;

Db 5 kwcftlnyssaaredflalkeedvhyvsvvgdevapatgkhlglgylslkrlrlg 64  
 QY 16 RWFTLNPNSEDERKKIRELPISL-FDYFVIGEGNEEGRTPHLQGFANFVKQTEN 74

Db 65 kky-gsrahweiaargdeqnrnycske-tlilvlelgtvpvpgsk 105  
 QY 75 W-YLGARCHIEKAKGTDQONKEYCSKREGNLLIECGAPRSOGQR 116

RESULT 3  
 ID W87461 standard; Protein; 286 AA.  
 AC W87461;  
 DT 10-FEB-1999 (first entry)  
 DE BBTV DNA II V2 C1 amino acid sequence.  
 KW Banana bunchy top virus; BBTV; PCR technique; plant; tissue; banana;  
 MU Musa; phloem damage; aphid; nucleic acid detection.  
 OS Banana bunchy top virus.  
 PN US5846705-A.  
 PD 08-DEC-1998.  
 PF 06-APR-1995; 418071.  
 PR 06-APR-1995; US-418071.  
 PA (BIOT-) DEV CENT BIOTECHNOLOGY.

PI Soong T, Wu R, You L;  
 DR WPI; 99-059037/05.  
 PT Nucleic acids having banana bunchy top virus component sequences -  
 PT used to design primers for use in polymerase chain reaction  
 PT detection of the virus  
 PS Disclosure; Columns 27-30; 27pp; English.  
 CC This represents the amino acid sequence of a banana bunchy top virus  
 CC (BBTV) DNA II V2 C1 protein. The invention provides nucleic acid  
 CC sequences associated with BBTV that can be used in a PCR technique for  
 CC detecting BBTV. The nucleic acid sequences (V71830 to V71833) are used as  
 CC the basis for the construction of PCR primers, to detect BBTV infection.  
 CC The PCR technique is used for detecting BBTV in plant tissues (preferably  
 CC banana, especially Musa species). The virus, one of the most important  
 CC banana species viruses, causes phloem damage and is transmitted by  
 CC aphids. PCR detection gives accurate, reliable and specific determination  
 CC of absence or presence of the virus.  
 SQ Sequence 286 AA;

Query Match 5.98; Score 139; DB 38; Length 286;  
 Best Local Similarity 29.88; Pred. No. 4.05e-03;  
 Matches 25; Conservative 28; Mismatches 23; Indels 8; Gaps 8;

Db 8 wmtfnnpt-tlpv-mrd-eik-ykvqv-drg-qeg-trhvggyvnmkrsslkqmrfg 60  
 QY 17 WVFTLNPNSEDERKKIRELPISLFDYFVIGEGNEEGRTPHLQGFANFVKQTEN 76

Db 61 fpga-hlekrkgsqearysckme 83  
 QY 77 LGARCHIEKAKGTDQONKEYCSKE 100

RESULT 4  
 ID W54071 standard; Protein; 286 AA.  
 AC W54071;  
 DT 11-AUG-1998 (first entry)  
 DE Banana bunchy top virus Component 1 protein.  
 KW BBTV; probe; diagnostic primer; component 1.  
 OS Banana bunchy top virus.  
 PN US5756708-A.  
 PD 26-MAY-1998.  
 PF 24-FEB-1994; 202186.  
 PR 24-FEB-1994; US-202186.  
 PA (UQO-) UNIV QUEENSLAND TECHNOLOGY.  
 PI Burns TM, Dale JL, Harding RM, Karan M;  
 DR WPI; 98-321636/28.  
 DR N-PSDB; V24077.  
 PT Isolated Banana bunchy top virus DNA - useful as diagnostic probes  
 PT and primers and for producing virus-resistant plants  
 PS Example 1; Fig 11; 59pp; English.  
 CC This sequence represents the protein encoded by the component 1 of  
 CC the Banana bunchy top virus (BBTV). The DNA molecules can be used as  
 CC diagnostic probes or primers or can be inserted into plants or other  
 CC organisms e.g. to produce virus-resistant plants or to act as promoters,  
 CC enhancers or termination signals.  
 SQ Sequence 286 AA;

Query Match 5.88; Score 136; DB 31; Length 286;  
 Best Local Similarity 31.08; Pred. No. 7.38e-03;  
 Matches 26; Conservative 27; Mismatches 23; Indels 8; Gaps 8;

Db 8 wmtfnnpt-tlpv-mrd-eik-ymvqv-erg-qeg-trhvggyvnmkrsslkqmrfg 60  
 QY 17 WVFTLNPNSEDERKKIRELPISLFDYFVIGEGNEEGRTPHLQGFANFVKQTEN 76

Db 61 fpga-hlekrkgsqearysckme 83  
 QY 77 LGARCHIEKAKGTDQONKEYCSKE 100

RESULT 5  
 ID W02157 standard; Protein; 611 AA.  
 AC W02157;  
 DT 14-JAN-1997 (first entry)

DE Periplasmic Beta-N-acetylglucosaminidase.  
 KW Periplasmic chitinase; periplasmic beta-N-acetylglucosaminidase;  
 KW Beta-N-acetylglucosaminidase; chitin; oligosaccharide; catabolic;  
 KW catabolism.  
 OS Vibrio furnissii.  
 PN W09625424-A1.  
 PD 22-AUG-1996. U02332.  
 PR 13-FEB-1996; US-386727.  
 PR 13-FEB-1995; US-386727.  
 PA (UJO) UNIV JOHNS HOPKINS.  
 PI Bassler B, Chitlaru E, Keyhani N, Roseman S, Rowe C;  
 PI Yu C;  
 DR WPI: 96-393335/39.  
 DR N-PSDB: T36388.  
 DR Chitin biosynthetic enzymes end I, exo I and exo II - are  
 PT periplasmic chitinase(s), periplasmic beta-GlcNAcidase(s) and  
 PT aryl beta-N-acetylglucosaminidase(s), respectively  
 PS Claim 10; Page 73-75; 101pp; English.  
 CC Periplasmic chitinase (W02156), periplasmic  
 CC Beta-N-acetylglucosaminidase (W02157) and aryl  
 CC Beta-N-acetylglucosaminidase (W02158) can be used to produce chitin  
 CC oligosaccharides with the structure (GlcNAc)<sub>n</sub> where n is 2 or  
 CC higher, by contacting them with soluble chitin. The enzymes are  
 CC encoded by the genes endI, exoI and exoII respectively. They are  
 CC all genes involved in the catabolic pathway of chitin.  
 SQ Sequence 611 AA;

Query Match 4.1%; Score 97; DB 19; Length 611;  
 Best Local Similarity 25.3%; Pred. No. 1.22e+01;  
 Matches 19; Conservative 20; Mismatches 31; Indels 5; Gaps 5;  
 Db 484 weeaqgdkvskdtviyswseqaln-carggfdvllpqgftldiaqdyapeepgvd 542  
 QY 203 WD-GYGEKVVVIDFYGWLPPWDLRLCDRPLVTKGTVFPLARSILITSNOTPLE 261  
 Db 543 w-agvt-p-lerayr 554  
 QY 262 WYSSTAVPAVEALYR 276

RESULT 6  
 ID W85599 standard; protein; 611 AA.  
 AC W85599;  
 DT 02-MAR-1999 (first entry)  
 DT 08-MAR-1999 (revised)  
 DE Hexosaminidase enzyme.  
 KW Hexosaminidase; enzyme; laundry; cleaning agent; hydrolysis;  
 KW anti-microbial; detergent; surfactant.  
 OS Unidentified.  
 PN W09850512-A1.  
 PD 12-NOV-1998.  
 PF 05-MAY-1998; U09125.  
 PR 19-AUG-1997; US-056132.  
 PR 06-MAY-1997; US-045756.  
 PA (PROC) PROCTER & GAMBLE CO.  
 PI Convents AC, Moese RL, Wolff AM;  
 DR WPI: 99-024116/02.  
 DR N-PSDB: V83129.  
 PT Laundry and cleaning compositions containing hexosaminidase - to  
 PT provide antimicrobial activity and remove biofilm  
 PS Claim 2; Page 38-39; 64pp; English.  
 CC Novel hexosaminidase enzymes (W85599-605) can be used in  
 CC combination in an aqueous laundry or cleaning product. The cleaning  
 CC product is used especially used to launder fabrics and to clean  
 CC dishes and tableware, particularly in an automatic dishwasher, but  
 CC may also be used generally as hard surface cleaner. The cleaning  
 CC product imparts antimicrobial activity and/or eliminates biofilm,  
 CC the hexosaminidases having a minimum inhibitory concentration of  
 CC less than about 0.125% but more preferably less than about 0.025%.  
 CC (NB: entry was revised to change incorrect cross references in  
 CC comments field).  
 SQ Sequence 611 AA;

Query Match 4.1%; Score 97; DB 39; Length 611;  
 Best Local Similarity 25.3%; Pred. No. 1.22e+01;  
 Matches 19; Conservative 20; Mismatches 31; Indels 5; Gaps 5;  
 Db 484 weeaqgdkvskdtviyswseqaln-carggfdvllpqgftldiaqdyapeepgvd 542  
 QY 203 WD-GYGEKVVVIDFYGWLPPWDLRLCDRPLVTKGTVFPLARSILITSNOTPLE 261  
 Db 543 w-agvt-p-lerayr 554  
 QY 262 WYSSTAVPAVEALYR 276

RESULT 7  
 ID W62673 standard; protein; 165 AA.  
 AC W62673;  
 DT 09-NOV-1998 (first entry)  
 DE Streptococcus pneumoniae polypeptide.  
 KW Polypeptide; ORF; open reading frame; infection; bacterial;  
 KW Streptococcus; bacteremia; diagnosis; prophylaxis.  
 OS Streptococcus pneumoniae.  
 PN W09823631-A1.  
 PD 04-JUN-1998.  
 PF 24-NOV-1997; U21976.  
 PR 27-NOV-1990; US-031879.  
 PA (SMIK) SMITHKLINE BEECHAM CORP.  
 PA (SMIK) SMITHKLINE BEECHAM PLC.  
 PI Reid RH, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO,  
 DR WPI: 98-322654/26.  
 PT Streptococcus pneumoniae polynucleotides - useful for developing  
 PT products for diagnosis, prevention and treatment of infections e.g.  
 PT pneumonia, bacteremia, meningitis or endocarditis  
 PS Claim 5; Page 32; 181pp; English.  
 CC The sequence is that of a Streptococcal polypeptide.  
 CC The polypeptide can potentially be used for the diagnosis and  
 CC prevention of bacterial infections, especially SP infection.  
 CC It may be used for the treatment of diseases such as otitis media,  
 CC conjunctivitis, pneumonia, bacteremia, meningitis, sinusitis, pleural  
 CC empyema, endocarditis or infection of the cerebrospinal fluid.  
 SQ Sequence 165 AA;

Query Match 3.9%; Score 92; DB 34; Length 165;  
 Best Local Similarity 18.9%; Pred. No. 2.94e+01;  
 Matches 14; Conservative 27; Mismatches 29; Indels 4; Gaps 4;  
 Db 86 qpslkqhsvvdtlikittffliqadnqrfqdlilfqltkngsmariarr-rlngw 144  
 QY 115 QPSDLSTAVSTLLESGI-LVT-VAEQHPVTFVKNFRGLAELLKVGKMRDKNTVHFI 172  
 Db 145 isrpacs-tkwans 157  
 QY 173 VGPPGCCGCKSWAAN 186

RESULT 8  
 ID R47468 standard; protein; 820 AA.  
 AC R47468;  
 DT 03-AUG-1994 (first entry)  
 DE Branching enzyme of rice starch.  
 KW Rice; starch; transit peptide; pectin; cereal; aminopectin; seeds;  
 KW reverse transcriptase; plaques.  
 OS Cryza.sativa.  
 FH Key Location/Qualifiers  
 FT peptide 1.60 /label= transit\_peptide  
 FT J05317057-A.  
 PN 03-DEC-1993.  
 PD 30-MAR-1992; 102499.  
 PR 20-SEP-1991; JP-48617.  
 PA (MITS-) MITSUI GYOSAI SHOKUBUTU BIO KENKYUSHO KK.  
 DR WPI: 94-011022/02.  
 DR N-PSDB: Q54f74.



PT Gene CDNA for rice starch branching enzyme for varied amino  
 PT pectin in cereal - comprises structural gene specified by basic  
 PT sequence introduced in rice plant for improved taste, for DNA  
 PT fragment originated from rice genome contg. gene  
 PS Claim 1; Page 5-7; 21pp; Japanese.  
 CC The sequence shows a cDNA encoding a branching enzyme of rice  
 CC starch. The enzyme can be used to modify amylopectin content of  
 CC starch in cereal particles by introducing the basic sequence into  
 CC a rice plant. This process can be used to improve the taste of the  
 CC rice.  
 SQ Sequence 820 AA;

Query Match 3.9%; Score 91; DB 9; Length 820;  
 Best Local Similarity 40.6%; Pred. No. 3.50e+01;  
 Matches 13; Conservative 8; Mismatches 9; Indels 2; Gaps 2;

Db 89 khdfnyrkyrdqkcliekheggleefsky 120

QY 67 KQTFN-KVWYLGARCHIEKAKGT-DQONKEY 96

# RESULT

ID R53228 standard; Protein; 820 AA.

AC R53228;

DT 06-MAR-1995 (first entry)

DE Rice starch branching enzyme.

KW Rice starch branching enzyme; oryza sativa; amylopectin; albumen;

KW starch.

OS Oryza sativa.

PN J06098656-A.

PD 12-APR-1994.

PF 30-MAR-1992; 102500.

PR 30-MAR-1992; JP-102500.

PA (MITS-) MITSUI GYOSAI SHOKUBUTSU BIO KENKYUSHO.

DR WPI; 94-155835/19.

DR N-PSDB; Q62135.

PT Transgenic rice containing the rice starch branch family enzyme -

PT used to increase the amylopectin content of albumen

PS Claim 3; Page 6-8; 24pp; Japanese.

CC The introduction of the rice starch branch-forming enzyme gene into

CC a rice increase the activity of this enzyme in the plant, thereby

CC increasing the content of amylopectin in albumen starch and thus

CC enabling efficient mass production of various proteins.

SQ Sequence 820 AA;

Query Match 3.9%; Score 91; DB 11; Length 820;  
 Best Local Similarity 40.6%; Pred. No. 3.50e+01;  
 Matches 13; Conservative 8; Mismatches 9; Indels 2; Gaps 2;

Db 89 khdfnyrkyrdqkcliekheggleefsky 120

QY 67 KQTFN-KVWYLGARCHIEKAKGT-DQONKEY 96

# RESULT

ID R55273 standard; Protein; 1822 AA.

AC R55273;

DT 31-JAN-1995 (first entry)

DE Beta subunit of integrin cell surface receptor.

KW Integrin; alpha; beta; subunit; glycoprotein; heterodimer;

KW transmembrane; extracellular matrix; cell signalling; cytoskeleton;

KW behaviour; signal transduction; receptor.

OS Homo sapiens.

FH Key Location/Qualifiers

FT peptide 28..42

FT /note= "N-terminal peptide."

FT modified\_site 491

FT /note= "Potential N-linked glycosylation site."

FT modified\_site 617

FT /note= "Potential N-linked glycosylation site."

FT modified\_site 695

FT /note= "Potential N-linked glycosylation site."

FT domain 711..733

FT modified\_site 980 /note= "Putative transmembrane domain."  
 FT /note= "Potential N-linked glycosylation site."  
 FT modified\_site 1593  
 FT /note= "Potential N-linked glycosylation site."  
 PN US5320942-A.  
 PD 14-JUN-1994.  
 PR 19-FEB-1987; 016552.  
 PR 19-FEB-1987; US-016552.  
 PR 04-JAN-1989; US-293384.  
 PR 01-OCT-1990; US-591105.  
 PA (KAIJI/) KAJIJI S.  
 PA (QUAK/) QUARANTA V.  
 PI Kajiiji S.; Quaranta V;  
 DR WPI; 94-191533/23.  
 DR N-PSDB; Q65674.  
 PT Diagnosing presence of abnormal epithelial tissue in vitro -  
 PT utilises monoclonal antibodies to alpha6 beta4 cell surface protein  
 PS Example 5; Figure 9; 34pp; English.  
 CC Integrins are heterodimers comprised of alpha and beta subunits, that  
 CC are non-covalently associated transmembrane glycoproteins. 11 alpha  
 CC chains and 6 beta chains have been recognised in man. Each alpha  
 CC subunit tends to associate with only one type of beta subunit but  
 CC there are several exceptions. Integrins mediate (in part) the  
 CC interaction of cells with the extracellular matrix, forming a link  
 CC between the extracellular matrix and the cytoskeleton. They may  
 CC transmit signals from the extracellular to the intracellular  
 CC environment, affecting cell behaviour. This sequence is the beta4  
 CC subunit of an alpha6 beta4 integrin.  
 SQ Sequence 1822 AA;

Query Match 3.9%; Score 92; DB 10; Length 1822;  
 Best Local Similarity 28.8%; Pred. No. 2.94e+01;  
 Matches 17; Conservative 19; Mismatches 19; Indels 4; Gaps 4;

Db 1013 irrvldgksgvyrtdgtaggrdyipvegellfpggeawkelqkllleqvdsll 1071  
 QY 73 VKWYL-GARCHIE-KAR-GTDQONKEYCSKEGNLLIEGAPRSQGRSDLS-TAVSTILL 127

# RESULT

ID P60045 standard; Protein; 2150 AA.

AC P60045;

DT 30-JUL-1991 (first entry)

DE Sequence of viral proteins VP-1-VP4, P2A-P2C, P3A, VPg, protease

DE. Polymerase.

KW Picornavirus; therapy.

OS Rhinovirus HRV2.

FH Key Location/Qualifiers

FT protein 1..69

FT /label= VP4

FT protein 70..330

FT /label= VP2

FT protein 331..567

FT /label= VP3

FT protein 568..855

FT /label= VP1

FT protein 856..881

FT /label= ?

FT protein 882..991

FT /label= ? P2-A

FT protein 992..1086

FT /label= P2-B

FT protein 1087..1409

FT /label= P2-C

FT protein 1410..1486

FT /label= P3-A

FT protein 1487..1507

FT /label= VPg

FT protein 1508..1690

FT /label= Protease

FT protein 1691..2150

FT /label= Polymerase



## RESULT 14

ID P60066 standard; Protein; 2227 AA.  
 AC P60066;  
 DT 26-JUN-1991 (first entry)  
 DE Sequence of viral I434 polypeptide encoded by the complete  
 DE Nucleotide sequence of the HAV genome.  
 KW Diagnosis; vaccine; passive immunotherapy.  
 OS Hepatitis A virus.  
 FH Key Location/Qualifiers  
 FT region 1..245  
 FT /label= P1.1A  
 FT region 246..491  
 FT /label= 1B  
 FT region 492..836  
 FT /label= 1C  
 FT region 837..980  
 FT /label= P2.2A  
 FT region 981..1076  
 FT /label= 2B  
 FT region 1077..1422  
 FT /label= 2C  
 FT region 1423..1484  
 FT /label= P3.3A  
 FT region 1485..1507  
 FT /label= 3B  
 FT region 1508..1678  
 FT /label= 3C  
 FT region 1679..2227  
 FT /label= 3D  
 PN EP-199480-A.  
 PD 29-OCT-1986.  
 PR 03-APR-1986; 302465.  
 PR 03-APR-1985; US-719329.  
 PA (CHIR-) CHIRON CORP.  
 PI Dina D. Potter SJ, Vannest GA, Caput D;  
 DR WPI: 86-286213/44.  
 DR N-PSDB; N60080.  
 DT Hepatitis A virus nucleotide sequence and polypeptide - and use  
 PT in prodn. of vaccines and diagnostic probes  
 PS Claim 5; Fig 1; 18pp; English.  
 CC N60080 and oligonucleotide fragments are useful in detection of  
 CC hepatitis A virus; transformed hosts may be used for expression of  
 CC polypeptides and fragments useful in vaccines without risk of  
 CC infection by the virus or in prodn. of particles which are capable  
 CC of inducing immunocompetent B cells for passive immunotherapy. Pref.  
 CC epitope is derived from AAs 445-657 or 792-848 of the HAV  
 CC polypeptide sequence (P60066).  
 SQ Sequence 2227 AA;

Query Match 3.9%; Score 92; DB 3; Length 2227;  
 Best Local Similarity 46.7%; Pred. No. 2.94e+01;  
 Matches 14; Conservative 5; Mismatches 10; Indels 1; Gaps 1;  
 Db 1254 epeknlytkpvadsgydgsgqlvcilddi 1283  
 QY 189 NPETT-YWKPPKRWDCYHGEKVVVDDF 217

## RESULT 15

ID W34074 standard; Protein; 2227 AA.  
 AC W34074;  
 DT 27-APR-1998 (first entry)  
 DE Hepatitis A virus HM-175 protein sequence.  
 KW HAV; attenuation; 2C protein; 2C chimeric hepatitis A virus;  
 OS infection; vaccine.  
 FH Key Location/Qualifiers  
 FT Protein 1..23  
 FT /label= VP4  
 FT Protein 24..245  
 FT /label= VP2  
 FT Protein 246..491  
 FT /label= VP3

FT Protein 492..791  
 FT /label= VP1  
 FT Protein 792..980  
 FT /label= 2A  
 FT Protein 981..1087  
 FT /label= 2B  
 FT Protein 1088..1422  
 FT /label= 2C  
 FT Protein 1423..1496  
 FT /label= 3A  
 FT Protein 1497..1519  
 FT /label= 3B  
 FT Protein 1520..1738  
 FT /label= 3C  
 FT Protein 1739..2227  
 FT /label= 3D  
 PN W09740166-A2.  
 PD 30-OCT-1997.  
 PR 18-APR-1997; U06506.  
 PR 19-APR-1996; US-015642.  
 PA (USSH ) US SEC DEPT HEALTH..  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 PI Emerson SU, Purcell RH, Raychaudhuri G;  
 DR WPI: 97-535850/49.  
 DR N-PSDB; T93023.  
 DT Human attenuated HAV genome containing simian HAV 2C gene - useful  
 PT as vaccines against HAV infection  
 PS Disclosure; Fig 13A-D; 66pp; English.  
 CC This protein sequence is encoded by the human hepatitis A virus  
 CC (HAV) HM-175 wild-type genome (see T93023). Attenuated strain  
 CC HAV/7 is obtained by passage of HM-175 in African Green Monkey  
 CC kidney cells. A claimed DNA construct (1) comprises a genome of  
 CC HAV, where the genome is a human attenuated HAV genome in which a  
 CC region of the 2C gene has been replaced by a corresponding region  
 CC from a 2C gene of a simian AGM-27 HAV genome (see T93024). The  
 CC region of the 2C gene from AGM-27 contained in the construct  
 CC preferably encodes amino acids 120-328 of the 2C protein, amino  
 CC acids 1-121 or amino acids 1-328. Also claimed are: (1) an RNA  
 CC transcript of (1); (2) a cell transfected with (1) or the RNA  
 CC transcript of (1); (3) a HAV genome as above; (4) antibodies to the  
 CC HAV of (3); and (5) a host cell containing the HAV of (3). (1) or  
 CC its RNA transcript, can be used as a vaccine for preventing HAV in  
 CC a mammal. (1) or the RNA transcript can also be used to stimulate  
 CC the production of protective antibodies in the mammal.  
 SQ Sequence 2227 AA;

Query Match 3.9%; Score 92; DB 27; Length 2227;  
 Best Local Similarity 46.7%; Pred. No. 2.94e+01;  
 Matches 14; Conservative 5; Mismatches 10; Indels 1; Gaps 1;  
 Db 1254 epeknlytkpvadsgydgsgqlvcilddi 1283  
 QY 189 NPETT-YWKPPKRWDCYHGEKVVVDDF 217

Search completed: Wed Dec 22 10:11:49 1999  
 Job time : 26 secs.

\*\*\*\*\*  
 W P S R L H  
 (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
 Run on: Wed Dec 22 10:14:00 1999; Maspar time 6.62 seconds  
 Tabular output not generated. 565.408 Million cell updates/sec

Title: >US-09-209-961-3  
 Description: (1-314) from US09209961.pep  
 Perfect Score: 2358  
 Sequence: 1 MPKNGRSGPQPKRWFT.....GGQFVTLSPCPPEPVEINY 314

Scoring table: PAM 150  
 Gap 11

Searched: 122461 seqs, 11912985 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: a-issued  
 1:5A\_COMB 2:5B\_COMB 3:PCT9\_COMB 4:backfiles1

Statistics: Mean 32.212; Variance 134.292; scale 0.240

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	242	10.3	290	2	US-08-202-Sequence 26, Applicati	2.13e-12
2	197	8.4	286	2	US-08-418-Sequence 16, Applicati	2.67e-08
3	193	8.2	285	2	US-08-418-Sequence 17, Applicati	6.10e-08
4	139	5.9	286	2	US-08-418-Sequence 18, Applicati	3.02e-03
5	136	5.8	286	2	US-08-202-Sequence 25, Applicati	5.38e-03
6	97	4.1	611	2	US-08-386-Sequence 4, Applicatio	7.06e+00
7	89	3.8	445	2	US-08-630-Sequence 4, Applicatio	2.78e+01
8	89	3.8	445	2	US-08-838-Sequence 4, Applicatio	2.78e+01
9	89	3.8	456	3	PCT-US95-1-Sequence 2, Applicatio	2.78e+01
10	89	3.8	456	1	US-08-349-Sequence 2, Applicatio	2.78e+01
11	90	3.8	1463	1	US-08-157-Sequence 3, Applicatio	2.35e+01
12	88	3.7	445	2	US-08-630-Sequence 6, Applicatio	3.29e+01
13	88	3.7	445	1	US-08-838-Sequence 6, Applicatio	3.29e+01
14	88	3.7	455	1	US-08-349-Sequence 4, Applicatio	3.29e+01
15	88	3.7	456	3	PCT-US95-1-Sequence 4, Applicatio	3.29e+01
16	88	3.7	533	1	US-08-488-Sequence 6, Applicatio	3.29e+01
17	86	3.6	346	2	US-08-744-Sequence 2, Applicatio	4.60e+01
18	86	3.6	428	4	5179023-4 Patent No. 5179023	4.60e+01
19	86	3.6	429	1	US-08-261-Sequence 7, Applicatio	4.60e+01
20	86	3.6	429	1	US-07-983-Sequence 2, Applicatio	4.60e+01
21	86	3.6	429	1	US-07-602-Sequence 3, Applicatio	4.60e+01
22	86	3.6	429	1	US-07-602-Sequence 2, Applicatio	4.60e+01
23	86	3.6	429	1	US-08-261-Sequence 9, Applicatio	4.60e+01

24	86	3.6	429	1	US-08-261-Sequence 3, Applicatio	4.60e+01
25	83	3.5	627	2	US-08-700-Sequence 6, Applicatio	7.55e+01
26	83	3.5	627	2	US-08-466-Sequence 6, Applicatio	7.55e+01
27	83	3.5	1724	1	US-08-325-Sequence 3, Applicatio	7.55e+01
28	81	3.4	310	3	PCT-US95-0-Sequence 42, Applicati	1.05e+02
29	81	3.4	310	3	PCT-US95-0-Sequence 42, Applicati	1.05e+02
30	81	3.4	310	2	US-08-337-Sequence 42, Applicati	1.05e+02
31	81	3.4	310	2	US-08-433-Sequence 42, Applicati	1.05e+02
32	80	3.4	325	1	US-08-233-Sequence 55, Applicati	1.23e+02
33	81	3.4	346	1	US-08-458-Sequence 2, Applicatio	1.05e+02
34	91	3.4	346	1	US-08-213-Sequence 2, Applicatio	1.05e+02
35	91	3.4	346	1	US-08-747-Sequence 2, Applicatio	1.05e+02
36	81	3.4	346	1	US-08-460-Sequence 6, Applicatio	1.05e+02
37	91	3.4	346	1	US-08-299-Sequence 6, Applicatio	1.05e+02
38	91	3.4	365	1	US-08-089-Sequence 5, Applicatio	1.05e+02
39	91	3.4	365	2	US-08-421-Sequence 5, Applicatio	1.05e+02
40	91	3.4	365	1	US-08-421-Sequence 5, Applicatio	1.05e+02
41	91	3.4	365	2	US-08-851-Sequence 4, Applicatio	1.05e+02
42	80	3.4	480	2	US-08-833-Sequence 49, Applicati	1.23e+02
43	80	3.4	480	1	US-08-306-Sequence 49, Applicati	1.23e+02
44	80	3.4	480	1	US-08-476-Sequence 49, Applicati	1.23e+02
45	80	3.4	507	1	US-08-457-Sequence 23, Applicati	1.23e+02

## ALIGNMENTS

RESULT 1  
 ID US-08-202-186-26, STANDARD; PRT; 290 AA.

XX XXXXXX

Sequence 26, Application US/08202186

Sequence 26, Application US/08202186

Patent No. 5756708

GENERAL INFORMATION:

APPLICANT: KARAN, Mirko

APPLICANT: BURNS, Thomas M.

APPLICANT: DALE, James L.

APPLICANT: HARDING, Robert M.

TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus

NUMBER OF SEQUENCES: 60

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner

STREET: 3000 K Street, N.W.

CITY: Washington, D.C.

COUNTRY: USA

ZIP: 20007-5109

COMPUTER: READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/202,186

FILING DATE: 24-FEB-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: JEFFERY, Donald D.

REGISTRATION NUMBER: 19,980

REFERENCE/DOCKET NUMBER: 71611/102 FIVE

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202 672 5300

TELEFAX: 202 672 5399

TELEX: 904136

INFORMATION FOR SEQ ID NO: 26:

SEQUENCE CHARACTERISTICS:

LENGTH: 290 amino acids

TYPE: amino acid

TOPOLOGY: linear

SEQUENCE 290 AA: 33497 MW; 430051 CN;

Query Match 10.3%; Score 242; DB 2; Length 290;  
Best Local Similarity 38.2%; Pred. No. 2.13e-12;  
Matches 42; Conservative 26; Mismatches 36; Indels 6; Gaps 6;

DB 6 RWCFTLNTVETEEAANVVRISLNLV-YAIVGVDEVPSTGQRHLQGFTHLKTGRRLQG 64  
QY 15 KRWFTLNPSEDERKKI-REL-PISLFDYFVGEENEGRTPHLQGFANFVKKQTFNK 72  
DB 65 LKTVLGNDRHLIPLTRGSDQNRDYCSKE-RVLEHGVTPRPGVKRPLA 113  
QY 73 VKWYLGAR-CHIEKAKGTDOQNKEYCSKEGNLLIEGAPRSQG-QRSDLS 120

RESULT 2  
ID US-08-418-071-16 STANDARD; PRT; 286 AA.

XX AC xxxxxx

XX AC

XX DT

XX DT

XX DE

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

Sequence 16, Application US/08418071

Sequence 16, Application US/08418071

Patent No. 5846705

GENERAL INFORMATION:

APPLICANT: Wu, Rey-Yuh

APPLICANT: You, Li-Ru

APPLICANT: Soong, Tai-Seng

TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA

TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS AND METHOD FOR DET

TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS

NUMBER OF SEQUENCES: 18

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Neave

STREET: 1251 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: United States

ZIP: 10020

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Word Perfect 5.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/418,071

FILING DATE: 06-APR-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Haley Jr., James F.

REGISTRATION NUMBER: 27,794

REFERENCE/DOCKET NUMBER: DCB-1

TELEPHONE: (212) 596-9000

TELEFAX: (212) 596-9090

INFORMATION FOR SEQ ID NO: 16:

SEQUENCE CHARACTERISTICS:

LENGTH: 286 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

DESCRIPTION: /desc="DNA I-VI"

SEQUENCE 286 AA; 33108 MW; 434824 CN;

Query Match 8.4%; Score 197; DB 2; Length 286;

Best Local Similarity 36.5%; Pred. No. 2.67e-08;

Matches 42; Conservative 21; Mismatches 47; Indels 5; Gaps 5;

DB 4 PSLKWCFTLNTYSSAARENFSLKKEEDVHYAVGVDEVPATGKHQGLYSLKKRRLG 63

QY 13 PKRWFTLNPSEDERKKIRELPISL-FDYFVGEENEGRTPHLQGFANFVKKOTFN 71

DB 64 GLKKKY-GSRAHWEIARGTDENSKYCSK-CTLILELGFVPVWNSNKRKISEMYA 116  
QY 72 KVKW-YLGARCHIEKAKGTDOQNKEYCSKEGNLLIEGAPRSQG-QRSDLS 124

RESULT 3  
ID US-08-418-071-17 STANDARD; PRT; 285 AA.

XX AC xxxxxx

XX AC

XX DT

XX DT

XX DE

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

Sequence 17, Application US/08418071

Sequence 17, Application US/08418071

Patent No. 5846705

GENERAL INFORMATION:

APPLICANT: Wu, Rey-Yuh

APPLICANT: You, Li-Ru

APPLICANT: Soong, Tai-Seng

TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA

TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS

TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS

NUMBER OF SEQUENCES: 18

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Neave

STREET: 1251 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: United States

ZIP: 10020

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Word Perfect 5.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/418,071

FILING DATE: 06-APR-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Haley Jr., James F.

REGISTRATION NUMBER: 27,794

REFERENCE/DOCKET NUMBER: DCB-1

TELEPHONE: (212) 596-9000

TELEFAX: (212) 596-9090

INFORMATION FOR SEQ ID NO: 17:

SEQUENCE CHARACTERISTICS:

LENGTH: 285 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

DESCRIPTION: /desc="DNA II V2 (D2)"

SEQUENCE 285 AA; 32701 MW; 420559 CN;

Query Match 8.2%; Score 193; DB 2; Length 285;

Best Local Similarity 37.9%; Pred. No. 6.10e-08;

Matches 39; Conservative 18; Mismatches 42; Indels 4; Gaps 4;

DB 5 KWCFTLNTYSSAAREDFLALKKEEDVHYVGVDEVPATGKHGGLYSLKKRIRLGLK 64

QY 16 RWFTLNPSEDERKKIRELPISL-FDYFVGEENEGRTPHLQGFANFVKKOTFNKVK 74

DB 65 KKY-GSRAHWEIAGSDQNRRCYSKE-TLVLELGTVPVPGSK 105

QY 75 W-YLGARCHIEKAKGTDOQNKEYCSKEGNLLIEGAPRSQQR 116

RESULT 4

ID US-08-418-071-18 STANDARD; PRT; 286 AA.

XX AC xxxxxx



CC STATE: D.C.  
CC COUNTRY: USA  
CC ZIP: 20005  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/386,727  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: HOBBS, ANN S.  
CC REGISTRATION NUMBER: 36,830  
CC REFERENCE/DOCKET NUMBER: 4130/206916  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 202-861-3000  
CC TELEFAX: 202-822-0944  
CC TELEX: 6714627 CUSH  
CC INFORMATION FOR SEQ ID NO: 4:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 611 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 611 AA; 69501 MW; 1856094 CN;

Query Match 4.1%; Score 97; DB 2; Length 611;  
Best Local Similarity 25.3%; Pred. No. 7.06e+00;  
Matches 19; Conservative 20; Mismatches 31; Indels 5; Gaps 5;  
Db 484 WEEAAGDKVSKDVIYISWLSQAALN-CARQGFVDVILQPGQFTYLDIAQDYAPEEPGVD 542  
QY 203 WD-GYHGEKVVIDDFYGLPWDLRLCDRYPLTVTKGTVPFLARSILITSNQIPPLE 261  
Db 543 W-AGVT-P-LERAYR 554  
QY 262 WYSTAVPAVEALYR 276

RESULT 7  
ID US-08-630-118A-4 STANDARD; PRT; 445 AA.  
XX  
AC xxxxxx

## Sequence 4, Application US/08630118A

Sequence 4, Application US/08630118A  
Patent No. 5919901  
GENERAL INFORMATION:  
CC APPLICANT: Hu Ph.D., Yinghe  
CC APPLICANT: McCaleb Ph.D., Michael L.  
CC APPLICANT: Bloomquist Ph.D., Brian T.  
CC APPLICANT: Flores-Riveros Ph.D., Jaime R.  
CC APPLICANT: Cornfield Ph.D., Linda J.  
CC TITLE OF INVENTION: Neuropeptide Y Receptor and Nucleic Acid  
CC TITLE OF INVENTION: Sequences  
CC NUMBER OF SEQUENCES: 8  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff  
CC STREET: 300 South Wacker Drive, 32nd Floor  
CC CITY: Chicago  
CC STATE: IL  
CC COUNTRY: USA  
CC ZIP: 60606  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/386,727  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Greenfield Ph.D., Michael S.  
CC REGISTRATION NUMBER: 37,147  
CC REFERENCE/DOCKET NUMBER: 96,149/WH 405  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (312)715-1000  
CC TELEFAX: (312)715-1234  
CC INFORMATION FOR SEQ ID NO: 4:

CC SOFTWARE: PatentIn Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/630,118A  
CC FILING DATE: April 8, 1996  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Greenfield Ph.D., Michael S.  
CC REGISTRATION NUMBER: 37,147  
CC REFERENCE/DOCKET NUMBER: 96,149/WH 405  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (312)715-1000  
CC TELEFAX: (312)715-1234  
CC INFORMATION FOR SEQ ID NO: 4:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 445 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 445 AA; 50408 MW; 1091793 CN;

Query Match 3.8%; Score 89; DB 2; Length 445;  
Best Local Similarity 39.4%; Pred. No. 2.78e+01;  
Matches 13; Conservative 10; Mismatches 10; Indels 0; Gaps 0;  
Db 49 TFSVLLGFMGNLLILMAVKKRNOKTTFVNFILIG 81  
QY 142 TFSVLLGFMGNLLILMAVKKRNOKTTFVNFILIG 174

RESULT 8  
ID US-08-838-399-4 STANDARD; PRT; 445 AA.  
XX  
AC xxxxxx

## Sequence 4, Application US/08838399

Sequence 4, Application US/08838399  
Patent No. 5965392  
GENERAL INFORMATION:  
CC APPLICANT: Hu Ph.D., Yinghe  
CC APPLICANT: McCaleb Ph.D., Michael L.  
CC APPLICANT: Bloomquist Ph.D., Brian T.  
CC APPLICANT: Flores-Riveros Ph.D., Jaime R.  
CC APPLICANT: Cornfield Ph.D., Linda J.  
CC TITLE OF INVENTION: Neuropeptide Y Receptor and Nucleic Acid  
CC TITLE OF INVENTION: Sequences  
CC NUMBER OF SEQUENCES: 8  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff  
CC STREET: 300 South Wacker Drive  
CC CITY: Chicago  
CC STATE: IL  
CC COUNTRY: USA  
CC ZIP: 60606  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/838,399  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Greenfield Ph.D., Michael S.  
CC REGISTRATION NUMBER: 37,147  
CC REFERENCE/DOCKET NUMBER: 96,149/WH 405  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (312)715-1000  
CC TELEFAX: (312)715-1234  
CC INFORMATION FOR SEQ ID NO: 4:









Tue Jan 4 11:52:26 2000

US-09-209-961-3.rai

Page 8

Search completed: Wed Dec 22 10:14:10 1999 ...  
Job time : 10 secs.

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

```
>US-09-209-961-3
Title:
Description: (1-314) from US09209961.pap
Perfect Score: 2358
Sequence: 1 MFPSKNGRSGQPQHKRWYTF.....GGQFVTLSPCPPEFYEINY 314
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Query		Score	Match	Length	DB	ID	Description	Pred. No.
1	242	10.3	290	2	AJ6353	ORF1 protein - coconu	2.95e-26		
2	136	5.8	286	2	Q19630	hypothetical 33.6K pr	1.09e-06		
3	130	5.5	3085	3	T00327	polyprotein - infecti	1.04e-05		
4	119	5.0	148	2	S04807	hypothetical protein,	5.62e-04		
5	116	4.9	295	2	D42452	Ci protein - tobacco	1.61e-03		
6	102	4.3	260	2	C71416	probable ABC transpor	1.78e-01		
7	99	4.2	1864	1	Q14557	genome polyprotein B	4.61e-01		
8	100	4.2	2157	1	GNNY1B	genome polyprotein -	3.36e-01		
9	97	4.1	244	2	JE0170	dnaj heat shock prote	8.61e-01		
10	96	4.1	273	2	S04125	chlorophyll a/b-bindi	1.17e+00		
11	96	4.1	417	2	A64380	phosphoglycerate kina	1.17e+00		
12	96	4.1	1777	3	T00490	nonstructural protein	1.17e+00		
13	95	4.0	94	2	H70654	hypothetical protein	1.59e+00		
14	95	4.0	286	2	S17696	chlorophyll a/b-bindi	1.59e+00		
15	95	4.0	415	2	S13627	ribulose-bisphosphate	1.59e+00		
16	94	4.0	503	1	ALRPP	alpha-amylase (EC 3.2	2.16e+00		
17	91	3.9	136	2	UQ0044	hypothetical 15.8K pr	5.28e+00		
18	91	3.9	159	2	J83299	sigma-E factor regula	5.28e+00		
19	91	3.9	211	2	S34274	ANG12 protein precurs	5.28e+00		
20	91	3.9	243	2	A25599	clathrin light chain	5.28e+00		
21	91	3.9	248	2	A31775	clathrin light chain	5.28e+00		
22	91	3.9	286	1	LRRY1A	clathrin light chain	5.28e+00		
23	92	3.9	313	2	H70777	probable cobD - Mvco	3.93e+00		

```

#accession J01960
##molecule_type mRNA
##residues 1-286 ##label HAR
KEYWORDS
FEATURE
180-188
SUMMARY
#region nucleotide-binding motif A (P-loop)
#length 286 #molecular-weight 33603 #checksum 3909

Query Match
Best Local Similarity 5.8%; Score 136; DB 2; Length 286;
Matches 26; Conservative 27; Mismatches 23; Indels 8; Gaps 8;

Db 8 WMTINNP-TLPV-WED-EIK-YMYQV-ERG-QEG-TRHVQGVEMKRSSLKQMRGF 60
QY 17 WFTLNPNSEDERKKIRLPISLFDYFIVGEEGNEGRTPHLOGFANFVKQTFNKVKY 76
Db 61 FPGA-HLEKRGKSGEARSYCMKE 83
QY 77 LGARCHIEKAKGTDQONKEYCSKE 100

RESULT 3
ENTRY
TITLE #type complete
ORGANISM polyprotein - infectious flacherie virus
DATE 29-Jan-1999 #sequence_revision 29-Jan-1999 #text_change
29-Jan-1999
ACCESSIONS T00327
REFERENCE Z14139
#authors Isawa, H.; Asano, S.; Sahara, K.; Iizuka, T.; Bando, H.
#journal Arch. Virol. (1998) 143:127-143
#title Analysis of genetic information of an insect picorna-like virus, infectious flacherie virus of silkworm: evidence for evolutionary relationships among insect, mammalian and plant picorna(-like) viruses.

#accession T00327
#status preliminary
#residues 1-3085 ##label ISA
#cross-references EMBL:AB000906; NID:d1185277; PID:d1026299
SUMMARY #length 3085 #molecular-weight 345801 #checksum 4939

Query Match
Best Local Similarity 5.5%; Score 130; DB 3; Length 3085;
Matches 12; Conservative 13; Mismatches 9; Indels 0; Gaps 0;

Db 1411 PKRFEHPNPVTRNPTSQMWNGYNGQPIVLYDDI 1444
QY 184 AANFANPETTYWRPKKWDGYHGERVVIDDF 217

RESULT 4
ENTRY
TITLE #type complete
ORGANISM hypothetical protein, 17.2K - maize streak virus (South African isolate)
DATE 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change
09-Sep-1997
ACCESSIONS S04807
REFERENCE S04804
#authors Lazarowitz, S.G.
#journal Nucleic Acids Res. (1988) 16:229-249
#title Infectivity and complete nucleotide sequence of the genome of a South African isolate of maize streak virus.
#cross-references MUID:88124198
#accession S04807
#status translation not shown
##molecule_type DNA
##residues 1-148 ##label LAZ
#cross-references EMBL:Y00514; NID:G59365; PID:G59369
SUMMARY #length 148 #molecular-weight 17214 #checksum 3101

Query Match
Best Local Similarity 5.0%; Score 119; DB 2; Length 148;
Matches 12; Conservative 13; Mismatches 9; Indels 0; Gaps 0;

#accession J01960
##molecule_type mRNA
##residues 1-286 ##label HAR
KEYWORDS
FEATURE
180-188
SUMMARY
#region nucleotide-binding motif A (P-loop)
#length 286 #molecular-weight 33603 #checksum 3909

Query Match
Best Local Similarity 5.8%; Score 136; DB 2; Length 286;
Matches 26; Conservative 27; Mismatches 23; Indels 8; Gaps 8;

Db 8 WMTINNP-TLPV-WED-EIK-YMYQV-ERG-QEG-TRHVQGVEMKRSSLKQMRGF 60
QY 17 WFTLNPNSEDERKKIRLPISLFDYFIVGEEGNEGRTPHLOGFANFVKQTFNKVKY 76
Db 61 FPGA-HLEKRGKSGEARSYCMKE 83
QY 77 LGARCHIEKAKGTDQONKEYCSKE 100

RESULT 3
ENTRY
TITLE #type complete
ORGANISM polyprotein - infectious flacherie virus
DATE 29-Jan-1999 #sequence_revision 29-Jan-1999 #text_change
29-Jan-1999
ACCESSIONS T00327
REFERENCE Z14139
#authors Isawa, H.; Asano, S.; Sahara, K.; Iizuka, T.; Bando, H.
#journal Arch. Virol. (1998) 143:127-143
#title Analysis of genetic information of an insect picorna-like virus, infectious flacherie virus of silkworm: evidence for evolutionary relationships among insect, mammalian and plant picorna(-like) viruses.

#accession T00327
#status preliminary
#residues 1-3085 ##label ISA
#cross-references EMBL:AB000906; NID:d1185277; PID:d1026299
SUMMARY #length 3085 #molecular-weight 345801 #checksum 4939

Query Match
Best Local Similarity 5.5%; Score 130; DB 3; Length 3085;
Matches 12; Conservative 13; Mismatches 9; Indels 0; Gaps 0;

Db 1411 PKRFEHPNPVTRNPTSQMWNGYNGQPIVLYDDI 1444
QY 184 AANFANPETTYWRPKKWDGYHGERVVIDDF 217

RESULT 4
ENTRY
TITLE #type complete
ORGANISM hypothetical protein, 17.2K - maize streak virus (South African isolate)
DATE 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change
09-Sep-1997
ACCESSIONS S04807
REFERENCE S04804
#authors Lazarowitz, S.G.
#journal Nucleic Acids Res. (1988) 16:229-249
#title Infectivity and complete nucleotide sequence of the genome of a South African isolate of maize streak virus.
#cross-references MUID:88124198
#accession S04807
#status translation not shown
##molecule_type DNA
##residues 1-148 ##label LAZ
#cross-references EMBL:Y00514; NID:G59365; PID:G59369
SUMMARY #length 148 #molecular-weight 17214 #checksum 3101

Query Match
Best Local Similarity 5.0%; Score 119; DB 2; Length 148;
Matches 12; Conservative 13; Mismatches 9; Indels 0; Gaps 0;

Matches 28; Conservative 28; Mismatches 37; Indels 12; Gaps 10;

Db 12 RSRKOSL-YIVGPTRTGKSTWARSGLVHN--YWNQVD--WSSYNEDAIYNTVDIPFKF 66
QY 163 RDWKTNVHFIVGPGCGSKWAANFANPETTYWPKPKKWDGYHGERVVIDDF-YGV 220
Db 67 CPCWKQLVG-CORDFIVNPKYKKKKKVOKSKPTIILANSDE-DW 109
QY 221 LP-WDDLRLCLDR-YPLTVK-TKGGTVFPLARSILITSNQTPLEW 262

RESULT 5
ENTRY
TITLE #type complete
ORGANISM Ci protein - tobacco yellow dwarf virus (strain Australia)
DATE 15-Jan-1993 #sequence_revision 15-Jan-1993 #text_change
17-Mar-1999
ACCESSIONS D42452
REFERENCE A42452
#authors Morris, B.A.M.; Richardson, K.A.; Haley, A.; Zhan, X.; Thomas, J.E.
#journal Virology (1992) 187:633-642
#title The nucleotide sequence of the infectious cloned DNA component of tobacco yellow dwarf virus reveals features of geminiviruses infecting monocotyledonous plants.
#cross-references MUID:92188538
#accession D42452
##molecule_type DNA
##residues 1-295 ##label MOR
#cross-references GB:M81103; NID:G335283; PID:G335287
SUMMARY #length 295 #molecular-weight 33651 #checksum 8742

Query Match
Best Local Similarity 4.9%; Score 116; DB 2; Length 295;
Matches 25; Conservative 29; Mismatches 26; Indels 7; Gaps 7;

Db 26 SSSAENLRDLWD-KLSRFAIFFTAIATELHODG-TPHLHCLIQDLKRSNRDPSFFDLE 83
QY 22 NNPSEDERKKIRLPISLFD-YFI-VGEEGNEGRTPHLOGFANFVKQTFNKVKY-L- 77
Db 84 GNHPNIOPAKNSEQ-VLEYISKDGNVI 109
QY 78 GARCHIEKAKGTDQONKEYCSKEGNLL 104

RESULT 6
ENTRY
TITLE #type complete
ORGANISM Probable ABC transporter - Arabidopsis thaliana
DATE 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change
05-Dec-1998
ACCESSIONS C71416
REFERENCE A71400
#authors Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirkse, W.; Van Staveren, M.; Stiekema, W.; Drost, L.; Ridley, P.; Hudson, S.A.; Patel, K.; Murphy, G.; Piffanelli, P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenegger, T.; Pohl, T.M.; Terryn, N.; Gielen, J.; Villarroel, R.; De Clerck, R.; Van Montagu, M.; Leclercq, A.; Auborg, S.; Gy, I.; Kreis, M.; Lao, N.; Kavanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.; Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech, P.; Douka, A.; Vouklatou, E.; Milioni, D.; Hatzopoulos, P.; Piravandi, E.; Obermaier, B.; Hilbert, H.; Duesterhoft, A.; Moeres, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Reckman, S.; Ansoorge, W.; Cooke, R.; Berger, C.; Delsen, M.; Voet, M.; Volckaert, G.; Newes, H.W.; Klosterman, S.; Schueller, C.; Chaiwatizis, N.
#journal Nature (1998) 391:485-488
#title Analysis of 1.9 Mb of contiguous sequence from chromosome 4

```

Query Match  
Best Local :

```

#cross-references MUID:98139904
#accession JE0170
##molecule_type mRNA
##residues 1-244 #label SIL
##cross-references GB:AF036875
GENETICS
#gene Mcd18
CLASSIFICATION #superfamily dnaJ amino-terminal homology
FEATURE
#domain dnaJ amino-terminal homology #label DNJ
#length 244 #molecular-weight 28206 #checksum 9987
SUMMARY
Query Match 4.1%; Score 97; DB 2; Length 244;
Best Local Similarity 21.5%; Pred. No. 8.61e-01;
Matches 14; Conservative 18; Mismatches 31; Indels 2; Gaps 2;
Db 79 FVELNEAYRVLRSRESRNDHDLH-SASPPKSSGSSTAEPRKYTOQTHTSWPPNAQYWAQ 137
QY 147 FRGLAELLKVGSKMQ-RDKMTNVHFTVGPCCGCKSKWAANFANPETIYWKPPKKNWDG 205
Db 138 FHSVR 142
QY 206 YHGEK 210
RESULT 10
ENTRY S04125 #type complete
TITLE Chlorophyll a/b-binding protein type III precursor - tomato
ORGANISM #formal_name lycopodium esculentum #common_name tomato
DATE 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change
21-Aug-1998
ACCESSIONS S04125
REFERENCE S04125
#authors Pichersky, E.; Brock, T.G.; Nguyen, D.; Hoffman, N.E.;
Piechulla, B.; Tanksley, S.D.; Green, B.R.
#journal Plant Mol. Biol. (1989) 12:257-270
#title A new member of the CAB gene family: structure, expression
and chromosomal location of cab-8, the tomato gene encoding
the type III chlorophyll a/b-binding polypeptide of
photosystem I.
#accession S04125
##molecule_type DNA
##residues 1-273 #label PIC
##cross-references EMBL:X15258
GENETICS
#gene cab-8
#introns 48/3; 69/2
CLASSIFICATION #superfamily chlorophyll a/b-binding protein
KEYWORDS chloroplast; light-harvesting complex; photosystem I;
transmembrane protein
SUMMARY #length 273 #molecular-weight 29293 #checksum 7937
Query Match 4.1%; Score 96; DB 2; Length 273;
Best Local Similarity 22.8%; Pred. No. 1.17e+00;
Matches 13; Conservative 22; Mismatches 16; Indels 6; Gaps 5;
Db 95 WLAYGEVIN--GRFAM-LGAAGATAPEILGKAGLI-POETALAWFOFGVVPACT-Y 146
QY 220 WLPWDDLLRLCDRYPLTVKTKGTVF-FLARSILITSNQPLEWYSTAVPAVEALY 275
RESULT 11
ENTRY A64380 #type complete
TITLE phosphoglycerate kinase (EC 2.7.2.3) - Methanococcus
ORGANISM jannaschii
#formal_name Methanococcus jannaschii
DATE 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change
10-Oct-1997
ACCESSIONS A64380
REFERENCE A64300
#authors Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann,
R.D.; Sutton, G.G.; Blake, J.A.; FitzGerald, L.M.; Clayton,
R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.;
Tombl, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.;
Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
Scott, J.L.; Geoghagen, N.S.M.; Weidman, J.F.; Fuhrmann,
J.D.; Nguyen, D.; Utterback, T.R.; Kelley, J.M.; Peterson,
J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts,
K.M.; Hurst, M.A.; Kaine, B.P.; Borodovsky, M.; Klenk, J.C.
#journal Science (1996) 273:1058-1073
#title Complete genome sequence of the methanogenic archaeon,
Methanococcus jannaschii.
#cross-references MUID:96337999
#accession A64380
#status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-417 #label BUL
##cross-references GB:U67512; GB:L7117; NID:gl591352; PID:gl5922299;
TIGR:MJ0641; PID:gl510724
GENETICS
#map_position REV571288-570035
CLASSIFICATION #superfamily phosphoglycerate kinase
KEYWORDS ATP: phosphotransferase
SUMMARY #length 417 #molecular-weight 46133 #checksum 9304
Query Match 4.1%; Score 96; DB 2; Length 417;
Best Local Similarity 30.8%; Pred. No. 1.17e+00;
Matches 16; Conservative 12; Mismatches 20; Indels 4; Gaps 4;
Db 130 KKWE-NITPKKQATNLKRLA-PLFDYF-VNDAFAAAHRAQPSLVGFSYIM 178
QY 15 KRWVTLNPNSEDERKKIRELPISLFDYFVIGEGNEGRT-PHLOGFANFV 65
RESULT 12
ENTRY T00490 #type complete
TITLE nonstructural protein precursor - himetobi P virus
ORGANISM #formal_name himetobi P virus
DATE 29-Jan-1999 #sequence_revision 29-Jan-1999 #text_change
29-Jan-1999
ACCESSIONS T00490
REFERENCE Z14156
#authors Nakashima, N.; Sasaki, J.
#submission submitted to the EMBL Data Library, August 1998
#description Complete nucleotide sequence of an insect picorna-like virus,
Himetobi P virus (Hivp).
#accession T00490
#status preliminary
##residues 1-1777 #label NAK
##cross-references EMBL:AB017037; NID:d1226972; PID:d1033516
SUMMARY #length 1777 #molecular-weight 201730 #checksum 5486
Query Match 4.1%; Score 96; DB 3; Length 1777;
Best Local Similarity 45.8%; Pred. No. 1.17e+00;
Matches 11; Conservative 5; Mismatches 8; Indels 0; Gaps 0;
Db 610 YSRASEQEFMDGYTGOLVTVFDDF 633
QY 194 YWKPPKKNWDGYHGERVVIDDF 217
RESULT 13
ENTRY H70654 #type complete
TITLE hypothetical protein Rv3851 - Mycobacterium tuberculosis
ORGANISM (strain H37RV)
#formal_name Mycobacterium tuberculosis
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
17-Jul-1998
ACCESSIONS H70654
REFERENCE A70500
#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry
III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.;

```

Search completed: Wed Dec 22 10:12:27 1999  
Job time : 20 secs.



\*\*\*\*\*  
W P S R L H  
\*\*\*\*\*  
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Dec 22 10:12:44 1999; MasPar time 11.43 Seconds  
Tabular output not generated. 776.734 Million cell updates/sec

Title: >US-09-209-961-3  
Description: (1-314) from US09209961.pep  
Perfect Score: 2358  
Sequence: 1 MPSKNGRSGPQPHRRVFT.....GGQFVTLSPPCPEPPEINY 314

Scoring table: PAM 150  
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot37  
1:swissprot

Statistics: Mean 47.377; Variance 72.055; scale 0.658

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	119	5.0	148	1 Y18K_MSVS	HYPOTHETICAL 17.2 KD P	3.62e-05
2	119	5.0	153	1 Y18K_MSVN	HYPOTHETICAL 17.7 KD P	3.62e-05
3	116	4.9	295	1 VALL_TYDA	ALI PROTEIN (C1 PROTEI	1.22e-04
4	99	4.2	1864	1 VGNB_CPMV	GENOME POLYPROTEIN B I	8.22e-02
5	100	4.2	2157	1 POLG_HRV1B	GENOME POLYPROTEIN [CO	5.72e-02
6	96	4.1	273	1 CB13_LYCES	CHLOROPHYLL A-B BINDIN	2.39e-01
7	96	4.1	417	1 PKG_METJA	PHOSPHOGLYCERATE KINA	2.39e-01
8	94	4.0	503	1 AMPY_RAT	ALPHA-AMYLASE PANCREAT	4.81e-01
9	95	4.0	2153	1 POLG_HRV16	GENOME POLYPROTEIN [CO	3.40e-01
10	91	3.9	136	1 Y16K_CSMV	HYPOTHETICAL 15.8 KD P	1.34e+00
11	91	3.9	159	1 RESE_ECOLI	SIGMA-E FACTOR REGULAT	1.34e+00
12	91	3.9	211	1 G12_ANOGA	PROTEIN G12 PRECURSOR	1.34e+00
13	91	3.9	243	1 CLCA_BOVIN	CLATHRIN LIGHT CHAIN A	1.34e+00
14	91	3.9	248	1 CLCA_HUMAN	CLATHRIN LIGHT CHAIN A	1.34e+00
15	91	3.9	248	1 CLCA_HUMAN	CLATHRIN LIGHT CHAIN A	1.34e+00
16	92	3.9	313	1 Y01H_MYCTU	HYPOTHETICAL 32.9 KD P	9.55e-01
17	91	3.9	376	1 MTC1_CITER	MODIFICATION METHYLASE	1.34e+00
18	91	3.9	590	1 ILVB_PORPU	ACETOLACTATE SYNTHASE	1.34e+00
19	92	3.9	812	1 FLVB_ECOLI	OUTER MEMBRANE USHER P	9.55e-01
20	91	3.9	820	1 ALGBA-ORISA	1,4-ALPHA-GLUCAN BRANC	1.34e+00
21	92	3.9	856	1 POLG_PVYO	GENOME POLYPROTEIN [CO	9.55e-01
22	93	3.9	879	1 GUNI_CLOTM	ENDOGLUCANASE I PRECUR	6.79e-01
23	91	3.9	1679	1 YMF9_YEAST	HYPOTHETICAL 187.1 KD	1.34e+00

24	92	3.9	1975	1 ITB4_HUMAN	INTEGRIN BETA-4 SUBUNI	9.55e-01
25	92	3.9	2150	1 POLG_HRV2	GENOME POLYPROTEIN [CO	6.79e-01
26	92	3.9	2226	1 POLG_HPAV4	GENOME POLYPROTEIN [CO	9.55e-01
27	92	3.9	2226	1 POLG_HPAV2	GENOME POLYPROTEIN [CO	9.55e-01
28	92	3.9	2226	1 POLG_HPAV8	GENOME POLYPROTEIN [CO	9.55e-01
29	92	3.9	2227	1 POLG_HPAVM	GENOME POLYPROTEIN [CO	9.55e-01
30	92	3.9	2227	1 POLG_HPAVL	GENOME POLYPROTEIN [CO	9.55e-01
31	92	3.9	2227	1 POLG_HPAVH	GENOME POLYPROTEIN [CO	9.55e-01
32	92	3.9	2230	1 POLG_HPAVS	GENOME POLYPROTEIN [CO	9.55e-01
33	99	3.8	456	1 NY5R_RAT	NEUROPEPTIDE Y RECEPT	2.60e+00
34	99	3.8	466	1 NY5R_MOUSE	NEUROPEPTIDE Y RECEPT	2.60e+00
35	99	3.8	781	1 GCSI_CABEL	PROBABLE MANNOSYL-OLIG	2.60e+00
36	99	3.8	1130	1 C2TA_HUMAN	MHC CLASS II TRANSACTI	1.87e+00
37	99	3.8	1866	1 VGNB_CPMV	GENOME POLYPROTEIN B I	2.60e+00
38	90	3.8	3859	1 RPOA_LELV	POLYPROTEIN (ORF1A	1.87e+00
39	90	3.8	4563	1 APB_HUMAN	APOLIPOPROTEIN B-100 P	1.87e+00
40	88	3.7	137	1 MGSB_BACSU	METHYLGLYOXAL SYNTHASE	3.61e+00
41	87	3.7	284	1 KDSA_ECOLI	2-DEHYDRO-3-DEOXYPHOS	5.00e+00
42	88	3.7	455	1 NY5R_HUMAN	NEUROPEPTIDE Y RECEPT	3.61e+00
43	88	3.7	497	1 KPYC_ARATH	PROBABLE PYRUVATE KINA	3.61e+00
44	88	3.7	718	1 STT3_YEAST	OLIGOSACCHARYL TRANSFE	3.61e+00
45	87	3.7	1699	1 POLN_LORDV	NON-STRUCTURAL POLYPRO	5.00e+00

ALIGNMENTS

RESULT 1  
ID Y18K\_MSVS STANDARD; PRT; 148 AA.  
AC P14989;  
DT 01-APR-1990 (REL. 14, CREATED)  
DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)  
D2 HYPOTHETICAL 17.2 KD PROTEIN.  
CS MAIZE STREAK VIRUS (SOUTH-AFRICAN ISOLATE) (MSV).  
OC VIRUSES; SSUVA VIRUSES; GEMINIVIRIDAE; MASTREVIUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 86124198.  
ZA LAZAROWITZ S.G.;  
RT "Infectivity and complete nucleotide sequence of the genome of a South African isolate of maize streak virus.";  
FL NUCLEIC ACIDS RES. 16:229-249(1988).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
CC  
CC EMBL; Y00514; G59369;  
DR PIR; S04807; S04897.  
KW HYPOTHETICAL PROTEIN.  
SQ SEQUENCE 148 AA; 17214 MW; E0133DAF CRC32;  
Query Match 5.0%; Score 119; DB 1; Length 148;  
Best Local: Similarity 26.7%; Pred. No. 3.62e-05;  
Matches 28; Conservative 28; Mismatches 37; Indels 12; Gaps 10;  
Db 12 RSRKSL-YIVGPTRTGKSTWARSIGVHN--YWNQNV--WSSYNEDAINVIDDIPKFK 66  
Oy 163 RDKKINVFHIVPPGCGKSKWAANFANPTTWTWPKNKNWDGYGHEKV-VVIDDF-YGW 220  
Db 67 CPCWKQLVG-CORDIVNPKYGGKKKKVKQKSKPTIILANSDE-DW 109  
Cy 221 LP-WDDLRLCDR-YPLTVK-TKGGIVPFLARSILITSNOTPLEW 262  
RESULT 2  
ID Y18K\_MSVN STANDARD; PRT; 153 AA.  
AC P14978;  
DT 01-APR-1990 (REL. 14, CREATED)

DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)  
DT 01-APR-1990 (REL. 14, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 17.7 KD PROTEIN.  
OS MAIZE STREAK VIRUS (NIGERIAN ISOLATE) (MSV).  
OC VIRUSES; SSDNA VIRUSES; GEMINIVIRIDAE; MASTREVIRUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 85126910.  
RA MULLINEAUX P.M., DONSON J., MORRIS-KRINITCH B.A.M., BOULTON M.I.,  
RT DAVIES J.W.;  
RA "The nucleotide sequence of maize streak virus DNA.";  
RL EMBO J. 3:3063-3068(1984).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X01633; G59362; -  
KW HYPOTHETICAL PROTEIN.  
SQ SEQUENCE 153 AA; 17768 MW; F1ED81B9 CRC32;  
  
Query Match 5.08; Score 119; DB 1; Length 153;  
Best Local Similarity 26.7%; Pred. No. 3.62e-05;  
Matches 28; Conservative 28; Mismatches 37; Indels 12; Gaps 10;  
  
DB 12 RSPKQSL-VIVGTRGKTSWASLGHVH--YWNQVDD--WSSYNNDAIYIVDDIPKFK 66  
QY 163 RDKTWNHIVGPPGCGKSKAANFANPETTYWKPKNKWDGHHGKV-VVIDDF-YGW 220  
DB 67 CPCKWLQV-CQRDFIVNPKYKKKKVQKSKPTIILANSDE-DW 109  
QY 221 LP-WDDLRLCDR-YPLTVK-TGGGVFPLARSILITSNTPLEW 262  
  
RESULT 3  
ID VAL1\_TYDVA STANDARD; PRT; 295 AA.  
AC P31617;  
DT 01-JUL-1993 (REL. 26, CREATED)  
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)  
DE 01-OCT-1993 (REL. 27, LAST ANNOTATION UPDATE)  
DE AL1 PROTEIN (C1 PROTEIN).  
GN C1.  
OS TOBACCO YELLOW DWARF VIRUS (STRAIN AUSTRALIA) (TYDV).  
OC VIRUSES; SSDNA VIRUSES; GEMINIVIRIDAE; MASTREVIRUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 92189538.  
RA MORRIS B.A.M., RICHARDSON K.A., HALEY A., ZHAN X., THOMAS J.E.;  
RT "The nucleotide sequence of the infectious cloned DNA component of  
RT tobacco yellow dwarf virus reveals features of geminiviruses  
RT infecting monocotyledonous plants.";  
RL VIROLOGY 187:633-642(1992).  
CC -!- FUNCTION: MAY BE INVOLVED IN REPLICATION.  
CC -!- SIMILARITY: BELONGS TO GEMINIVIRUSES AL1 PROTEIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M81103; G335287; -  
DR PIR; D42452; D42452.  
DR PFAM; PF00799; Geminivirus AL1; 1.  
SQ SEQUENCE 295 AA; 33651 MW; 520E4DB9 CRC32;  
  
Query Match 4.94; Score 116; DB 1; Length 295;

Best Local Similarity 28.7%; Pred. No. 1.22e-04;  
Matches 25; Conservative 29; Mismatches 26; Indels 7; Gaps 7;  
  
DB 26 SSSAENRDLWD-KLSRFAIFFIAIATELHQDG-TPHLHCLQLDKRSNRIDPSFFDLE 83  
QY 22 NNPSEDERKKIRELPISLFD-YFI-VGEEGNEGRTHLOQGFANFVKQTENKVKWY-L- 77  
DB 84 GNHPNIQAPNSEQ-VLEYISKDGNVI 109  
QY 78 GARCHIEKARGTDOONKEYCSKEGNLL 104  
  
RESULT 4  
ID VGNB\_RCMV STANDARD; PRT; 1864 AA.  
AC P35930;  
DT 01-JUN-1994 (REL. 29, CREATED)  
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)  
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
DE GENOME POLYPROTEIN B (CONTAINS: PROTEASE COFACTOR; MEMBRANE BINDING  
DE PROTEIN; VPG; PROTEASE (EC 3.4.22.-); RNA POLYMERASE (EC 2.7.7.48)).  
OS RED CLOVER MOTTLE VIRUS (RCMV).  
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; COMOVIRIDAE;  
CC COMOVIRUS.  
KA [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-S;  
RX MEDLINE; 93019077.  
RA SHANKS M., LOMONOSOFF G.P.;  
RT "The nucleotide sequence of red clover mottle virus bottom component  
RT RNA".  
RL J. GEN. VIROL. 73:2473-2477(1992).  
CC -!- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -!- SIMILARITY: TO OTHER COMOVIRUSES GENOME POLYPROTEIN B.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X64886; G61464; -  
DR PIR; JQ1657; JQ1657.  
DR PFAM; PF00680; RNA\_dep\_RNA\_pol; 1.  
DR PFAM; PF00910; RNA\_helicase; 1.  
KW POLYPROTEIN; TRANSMEMBRANE; HYDROLASE; PROTEASE; TRANSFERASE;  
KW RNA-DIRECTED RNA POLYMERASE; ATP-BINDING.  
FT CHAIN 1 315  
FT CHAIN 316 915  
FT CHAIN 916 943  
FT CHAIN 944 1151  
FT CHAIN 1152 1864  
FT NP\_BIND 483 490  
FT ACT\_SITE 983 983  
FT ACT\_SITE 1019 1019  
FT ACT\_SITE 1109 1109  
SQ SEQUENCE 1864 AA; 210255 MW; 8F34B5BD CRC32;  
  
Query Match 4.28; Score 99; DB 1; Length 1864;  
Best Local Similarity 40.7%; Pred. No. 8.22e-02;  
Matches 11; Conservative 6; Mismatches 10; Indels 0; Gaps 0;  
  
DB 509 DIVYSRNTTETHWSGYRROPVIWIDDF 535  
QY 191 ETTYWKPKKNKWDGHHGKRVWIDDF 217  
  
RESULT 5  
ID POLG\_HRVIB STANDARD; PRT; 2157 AA.  
AC P12916; Q89704; Q82106; Q82107; Q82108; Q82109; Q82110; Q82111;  
DT 01-OCT-1989 (REL. 12, CREATED)  
DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)





```

[1] SEQUENCE FROM N.A.
XP MEDJINE; 58219532.
RA ANDERSEN H.T., RICHARDSON K.A., HARBISON S.A., MORRIS B.A.M.;
RT "Nucleotide sequence of the geminivirus chloris striate mosaic
virus."
RL VIROLOGY 164:443-449(1988).
CC -1- WHILE THIS ORF DOES NOT HAVE AN ATG START CODON, THE SEQUENCE
CC SHOWS SIGNIFICANT HOMOLOGY, 69%, WITH THAT OF THE CORRESPONDING
CC ORF OF MAIZE STRIPY VIRUS. ALSO, THE CORRESPONDING ORFS OF TWO
CC OTHER GEMINIVIRUSES DO NOT HAVE AN ATG START CODON.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M20021; ; JUNT_ANNOTATED_CDS.
DR PIR; J00044; J00044.

```

```

SQ SEQUENCE 136 AA; 15762 MW; 5A406309 CRC32;

Query Match 3.9%; Score 91; DB 1; Length 136;
Best Local Similarity 26.2%; Pred. No. 1.34e+00;
Matches. 28; Conservative 24; Mismatches 44; Indels 11; Gaps 8;

Db 15 YICGPTTGTSWARS LGTH--YWHO SVN-FLEEWNCQAQFNLI DDIPKFPVCKHGLV 71
QY 171 FIVGFCGSKSAANFANPETTYKPKKNKWDGYHGE-KVVVIDDF-YGWL P-WDDLL 227
:b 72 G-SGYDLTVNPKYGGKKRIPGIPCIILVNEDE-DWLQSMSTQQVD 115
QY 228 RLCRYPLTVKTKGIVPFLARSI--LITSNQIPLEWYSSTAVPAVE 272

RESULT 11
ID RSEQ_ECOLI STANDARD; PRT; 159 AA.
AC P46187;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE SIGMA-55 FACTOR REGULATORY PROTEIN R5EC.
DE R5EC.
GN ESCHERICHIA COLI.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
FN 11

```

SEQUENCE FROM N.A.  
AC STRAIN=K12 / W3110;  
FA MISS:KAS D., RAINA S.;  
FL SUBMITTED (SEP-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RA CONNOLLY L., DE LAS PENAS A., GROSS C.A.;  
XL SUBMITTED (OCT-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RA NASHIMOTO H., SAITO N.;  
XL SUBMITTED (SEP-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE: 97426617  
RA BLATTNER F.R.; PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,  
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,  
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,  
RA MAU B., SHAO Y.;  
RT "The complete genome sequence of *Escherichia coli* K-12."

RT, "The complete genome sequence of *Escherichia coli* K-12."

RL SCIENCE 277:1453-1474(1997).  
 CC 1- FUNCTION: SEEMS TO MODULATE THE ACTIVITY OF RPOE (SIGMA-E).  
 CC 1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE  
 CC (POTENTIAL).  
 CC 1- SIMILARITY: TO H-INFLUENZAE H1589 AND H1850.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: U37089; G1045631; -  
 DR EMBL: U37455; G1050877; -  
 DR EMBL: D64044; G987645; -  
 DR EMBL: AE000343; G1788923; -  
 DR ECGENE; EG13178; RSEC.  
 KW TRANSMEMBRANE; INNER MEMBRANE.  
 FT TRANSMEM 80 100 POTENTIAL.  
 FT TRANSMEM 101 121 POTENTIAL.  
 SQ SEQUENCE 159 AA; 16639 MW; E1CDF35 CRC32;

Query Match 3.98; Score 91; DB 1; Length 159;  
 Best Local Similarity 35.5%; Pred. No. 1.34e+00;  
 Matches 11; Conservative 10; Mismatches 10; Indels 0; Gaps 0;

Db 51 IVVPCDEPLVPGKVGKVGIAEGSLSSALLV 81  
 QY 103 LLECCGAPRSQGRSDUSTAVTLLSGILV 133

RESULT 12  
 ID G12.ANOGA STANDARD; PRT; 211 AA.  
 AC Q17040;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DE PROTEIN G12 PRECURSOR (ANG12).  
 OS ANOPHELES GAMBIAE (AFRICAN MALARIA MOSQUITO).  
 OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;  
 OC PTERYGOTA; DIPTERA; NEMATOCERA; COLICOIDEA; CULICIDAE; ANOPHELES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SUAKOKO;  
 RA MUELLER H.M.; CRISANTI A.;  
 RL SUBMITTED (JUN-1993) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC 1- DEVELOPMENTAL STAGE: INDUCED IN THE MIDGUT OF FEMALE AFTER BLOOD  
 CC MEAL.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: Z22925; G311985; -  
 KW SIGNAL.  
 FT SIGNAL 1 19 POTENTIAL.  
 FT CHAIN 20 211 PROTEIN G12.  
 SQ SEQUENCE 211 AA; 23568 MW; 9ED0DC5 CRC32;

Query Match 3.98; Score 91; DB 1; Length 211;  
 Best Local Similarity 50.0%; Pred. No. 1.34e+00;  
 Matches 17; Conservative 4; Mismatches 13; Indels 0; Gaps 0;

Db 30 FDFPVGLPLNDLLDAMRYLLTDKEVQQTLLYL 63  
 QY 214 IDDFYGLPWDDLLRLUCDRYPLTVTKKGTVPFL 247

RESULT 13  
 ID CLCA\_BOVIN STANDARD; PRT; 243 AA.  
 AC P04973;  
 DT 13-AUG-1987 (REL. 05, CREATED)  
 DT 13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE CLATHRIN LIGHT CHAIN A (BRAIN AND LYMPHOCYTE LCA).  
 GN CLTA.  
 OS BOS TAURUS (BOVINE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC ARTIODACTYLA; RUMINANTIA; PECORA; BOVIDEA; BOVINAE; BOS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 87144634.  
 RA JACKSON A.P.; SLOW H.-F.; HOLMES N.; DRICKAMER K.; PARHAM P.;  
 RT "Clathrin light chains contain brain-specific insertion sequences and  
 FT a region of homology with intermediate filaments."  
 RL NATURE 326:154-159(1987).  
 CC 1- FUNCTION: CLATHRIN IS THE MAJOR PROTEIN OF THE POLYHEDRAL COAT OF  
 CC COATED PITS AND VESICLES.  
 CC 1- SUBUNIT: CLATHRIN COATS ARE FORMED FROM MOLECULES CONTAINING 3  
 CC HEAVY CHAINS AND 3 LIGHT CHAINS.  
 CC 1- SUBCELLULAR LOCATION: CYTOPLASMIC FACE OF COATED PITS AND  
 CC VESICLES.  
 CC 1- ALTERNATIVE PRODUCTS: MAMMALIAN CELLS CONTAIN 2 DISTINCT LIGHT  
 CC CHAIN GENES, LC(A) AND LC(B), WHICH ENCODE PRODUCTS 60% RELATED IN  
 CC SEQUENCE. TISSUE-SPECIFIC SPLICING PATTERNS GIVE RISE TO FOUR  
 CC DISTINCT FORMS OF CLATHRIN LIGHT CHAINS.  
 CC 1- THIS SEQUENCE IS IDENTICAL TO NON-BRAIN LC(A) EXCEPT THAT IT  
 CC CONTAINS A HYDROPHOBIC 30-RESIDUE, BRAIN-SPECIFIC INSERT.  
 CC THE NH2-TERMINAL PORTION OF THIS INSERT IS HOMOLOGOUS TO THE  
 CC BRAIN-SPECIFIC INSERT IN THE BRAIN LC(B) FORM.  
 CC 1- SIMILARITY: THERE IS A WEAK HOMOLOGY BETWEEN THE CENTRAL REGION OF  
 CC THE LIGHT CHAIN AND THE ALPHA-HELICAL PORTION OF VARIOUS IF  
 CC PROTEINS, PARTICULARLY CYTOKERATINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X04849; G508; -  
 DR EMBL: X04851; G510; -  
 DR PIR: A26599; A26599.  
 DR PIR: B26599; B26599.  
 DR PROSITE; PS00224; CLATHRIN LIGHT CHN.1; 1.  
 DR PROSITE; PS00581; CLATHRIN LIGHT CHN.2; 1.  
 DR PFAM; PF01086; Clathrin\_lg\_ch; 1.  
 DR MOD\_RES 1 1 BLOCKED.  
 FT CHAIN 95 157 INVOLVED IN BINDING CLATHRIN HEAVY  
 FT DOMAIN 158 187 MISSING (IN NON BRAIN FORMS).  
 FT VARSPIC 121 146 TO INTERMEDIATE FILAMENTS.  
 FT SIMILAR 14 14 P -> H (IN LYMPHOCYTE LCA).  
 FT CONFLICT 14 14  
 SQ SEQUENCE 243 AA; 26723 MW; 6C48315E CRC32;

Query Match 3.98; Score 91; DB 1; Length 243;  
 Best Local Similarity 39.5%; Pred. No. 1.34e+00;  
 Matches 15; Conservative 10; Mismatches 11; Indels 2; Gaps 2;

DB 206 WERVARLDCDFNPKSSK-QAKDVSRM-RSVLSLQAPL 241  
 QY 223 WDDLRLCDRYPLTVKTKGGTVFPFLARSILITSNOTPL 260

## RESULT 14

ID CLCA\_HUMAN STANDARD; PRT; 248 AA.  
 AC P09496;  
 DT 01-MAR-1989 (REL. 10, CREATED)  
 DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)  
 DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)  
 DE CLATHRIN LIGHT CHAIN A (BRAIN AND LYMPHOCYTE LCA).  
 GN CLTA.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 89034155.  
 RA JACKSON A.P.; PARHAM P.;  
 RT "Structure of human clathrin light chains. Conservation of light  
 chain polymorphism in three mammalian species.";  
 RL J. BIOL. CHEM. 263:16688-16695(1988).  
 CC -1- FUNCTION: CLATHRIN IS THE MAJOR PROTEIN OF THE POLYHEDRAL COAT OF  
 COATED PITS AND VESICLES.  
 CC -1- SUBUNIT: CLATHRIN COATS ARE FORMED FROM MOLECULES CONTAINING 3  
 HEAVY CHAINS AND 3 LIGHT CHAINS.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC FACE OF COATED PITS AND  
 VESICLES.  
 CC -1- ALTERNATIVE PRODUCTS: MAMMALIAN CELLS CONTAIN 2 DISTINCT LIGHT  
 CHAIN GENES, LC(A) AND LC(B), WHICH ENCODE PRODUCTS 60% RELATED IN  
 SEQUENCE. TISSUE-SPECIFIC SPLICING PATTERNS GIVE RISE TO FOUR  
 DISTINCT FORMS OF CLATHRIN LIGHT CHAINS.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
 DR EMBL; M20471; G179397;  
 DR EMBL; M20472; G307118;  
 DR PIR; A30752; A30752.  
 DR PIR; A30753; A30753.  
 DR PIR; A31775; A31775.  
 DR MIM; 118960;  
 DR PROSITE; PS00224; CLATHRIN\_LIGHT\_CHN\_1; 1.  
 DR PROSITE; PS00581; CLATHRIN\_LIGHT\_CHN\_2; 1.  
 DR PFAM; PF01086; Clathrin\_lg\_ch; 1.  
 KW COATED PITS; ALTERNATIVE SPLICING; CALCIUM-BINDING.  
 FT DOMAIN 100 162 INVOLVED IN BINDING CLATHRIN HEAVY  
 CHAIN.  
 FT VARSPLIC 163 192 MISSING (IN NON BRAIN FORMS).  
 FT SIMILAR 126 151 TO INTERMEDIATE FILAMENTS.  
 SQ SEQUENCE 248 AA; 27076 MW; 880CBA0B CRC32;

Query Match 3.9%; Score 91; DB 1; Length 248;  
 Best Local Similarity 39.5%; Pred. No. 1.34e+00;  
 Matches 15; Conservative 10; Mismatches 11; Indels 2; Gaps 2;

DB 211 WERVARLDCDFNPKSSK-QAKDVSRM-RSVLSLQAPL 246  
 QY 223 WDDLRLCDRYPLTVKTKGGTVFPFLARSILITSNOTPL 260

## RESULT 15

ID CLCA\_RAT STANDARD; PRT; 248 AA.  
 AC P08081;  
 DT 01-AUG-1988 (REL. 08, CREATED)  
 DT 01-AUG-1988 (REL. 08, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)

DE CLATHRIN LIGHT CHAIN A (BRAIN AND LYMPHOCYTE LCA).  
 GN CLTA.  
 OS RATTUS NORVEGICUS (RAT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 87178007.  
 RA KIRCHHAUSEN T.; SCARMATO P.; HARRISON S.C.; MONROE J.J.; CHOW E.P.;  
 RA MATALIANO R.J.; RAMACHANDRAN K.L.; SMART J.E.; AHN A.H.; BROSIUS J.;  
 RT "Clathrin light chains LCA and LCB are similar, polymorphic, and  
 share repeated heptad motifs.";  
 RL SCIENCE 236:320-324(1987).  
 CC -1- FUNCTION: CLATHRIN IS THE MAJOR PROTEIN OF THE POLYHEDRAL COAT OF  
 COATED PITS AND VESICLES.  
 CC -1- SUBUNIT: CLATHRIN COATS ARE FORMED FROM MOLECULES CONTAINING 3  
 HEAVY CHAINS AND 3 LIGHT CHAINS.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC FACE OF COATED PITS AND  
 VESICLES.  
 CC -1- ALTERNATIVE PRODUCTS: MAMMALIAN CELLS CONTAIN 2 DISTINCT LIGHT  
 CHAIN GENES, LC(A) AND LC(B), WHICH ENCODE PRODUCTS 60% RELATED IN  
 SEQUENCE. TISSUE-SPECIFIC SPLICING PATTERNS GIVE RISE TO FOUR  
 DISTINCT FORMS OF CLATHRIN LIGHT CHAINS.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
 DR EMBL; M15882; G203274;  
 DR EMBL; M19261; G203278;  
 DR EMBL; M19260; G203276;  
 DR PIR; A25994; LRRYAL.  
 DR PROSITE; PS00224; CLATHRIN\_LIGHT\_CHN\_1; 1.  
 DR PROSITE; PS00581; CLATHRIN\_LIGHT\_CHN\_2; 1.  
 DR PFAM; PF01086; Clathrin\_lg\_ch; 1.  
 KW COATED PITS; ALTERNATIVE SPLICING; CALCIUM-BINDING.  
 FT DOMAIN 100 162 INVOLVED IN BINDING CLATHRIN HEAVY  
 CHAIN.  
 FT VARSPLIC 163 192 MISSING (IN NON BRAIN FORMS).  
 FT SIMILAR 126 151 TO INTERMEDIATE FILAMENTS.  
 SQ SEQUENCE 248 AA; 26980 MW; 0E44BB68 CRC32;

Query Match 3.9%; Score 91; DB 1; Length 248;  
 Best Local Similarity 39.5%; Pred. No. 1.34e+00;  
 Matches 15; Conservative 10; Mismatches 11; Indels 2; Gaps 2;

DB 211 WERVARLDCDFNPKSSK-QAKDVSRM-RSVLSLQAPL 246  
 QY 223 WDDLRLCDRYPLTVKTKGGTVFPFLARSILITSNOTPL 260

Search completed: Wed Dec 22 10:12:58 1999  
 Job time : 14 secs.

\*\*\*\*\*  
WIPESH  
\*\*\*\*\*  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Wed Dec 22 10:15:55 1999; MasPar time 17.93 Seconds  
Tabular output not generated. 276.392 Million cell updates/sec

Title: >US-09-209-961-5  
Description: (1-233) from US09209961.pep  
Perfect Score: 1805  
Sequence: 1 MTYPRRYRRRRHRPSHLG.....RVMTYVQFREFNLKDPPLP 233  
Scoring table: PAM 150  
Gap 11

Searched: 170751 seqs, 21266508 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq35  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39

Statistics: Mean 33.499; Variance 147.654; scale 0.227

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	111	6.1	143	11	rAb Stage 1 VH constr	2.04e+00
2	108	6.0	143	27	Transplanted MAB HPI/	3.40e+00
3	108	6.0	143	13	Transplanted VH seque	3.40e+00
4	108	6.0	143	10	Humanised HPI/2 heavy	3.40e+00
5	108	6.0	143	36	Transplanted VH seque	3.40e+00
6	104	5.8	325	33	Amino acid sequence o	6.65e+00
7	103	5.7	124	11	SSE VH.	7.85e+00
8	103	5.7	143	11	STAW VH.	7.85e+00
9	103	5.7	143	11	Anti-VLA4 rAb Stage 2	7.85e+00
10	102	5.7	449	13	VPI protein.	9.27e+00
11	103	5.7	657	6	PMR-1 gene product.	7.85e+00
12	101	5.6	143	36	Humanised VH sequence	1.09e+01
13	101	5.6	143	10	Humanised HPI/2 heavy	1.09e+01
14	101	5.6	143	27	Humanised anti-VLA-4	1.09e+01
15	101	5.6	143	11	AS VH.	1.09e+01
16	101	5.6	143	11	Anti-VLA4 Ab AS VH se	1.09e+01

17	101	5.6	143	13	R70255	Anti-VLA-4 humanized	1.09e+01
18	99	5.5	143	11	R59931	KRS VH.	1.52e+01
19	99	5.5	449	4	R3829	Chicken anaemia virus	1.52e+01
20	98	5.4	449	16	R88497	VPI of chicken infect	1.79e+01
21	97	5.4	497	30	W49912	Mouse CLK serine/thre	2.11e+01
22	95	5.3	211	22	W06311	Hunt14.	2.91e+01
23	96	5.3	316	9	R45924	Modified thermolysin-	2.48e+01
24	95	5.3	402	26	R64200	Prostaglandin recepto	2.91e+01
25	96	5.3	1485	25	W17887	Photorhabdus luminesc	2.48e+01
26	96	5.3	1485	31	W56546	Toxin TscC, encoded b	2.48e+01
27	93	5.2	117	1	R80149	Biosynthetic antibody	4.02e+01
28	93	5.2	124	15	R82971	F85 humanized VH regi	4.02e+01
29	94	5.2	163	25	W32556	Mycobacterium tubercu	3.42e+01
30	94	5.2	163	35	W6296	Mycobacterium tubercu	3.42e+01
31	94	5.2	163	37	W81359	M. tuberculosis immun	3.42e+01
32	94	5.2	183	3	P60624	Sequence B encoded by	3.42e+01
33	93	5.2	316	1	P91634	Sequence encoding hea	4.02e+01
34	93	5.2	316	17	R97801	B.stearothermophilus	4.02e+01
35	93	5.2	316	9	R49837	Thermolysin-like meta	4.02e+01
36	93	5.2	316	26	W19127	Thermolysin protease	4.02e+01
37	93	5.2	316	8	R40972	N277R mutant neutral	4.02e+01
38	93	5.2	316	26	W19126	Thermolysin protease	4.02e+01
39	93	5.2	316	17	R97802	B.stearothermophilus	4.02e+01
40	93	5.2	316	17	R97795	B.stearothermophilus	4.02e+01
41	93	5.2	316	17	R97796	B.stearothermophilus	4.02e+01
42	93	5.2	316	8	R41515	Thermolysin like prot	4.02e+01
43	93	5.2	377	39	W89556	Triticum sp. cysteine	4.02e+01
44	94	5.2	518	39	W67616	A. nidulans phenylace	3.42e+01
45	93	5.2	552	2	R08203	Structural gene produ	4.02e+01

ALIGNMENTS

RESULT 1  
ID R59940 standard; Protein; 143 AA.  
AC R59940;  
DE 09-MAR-1995 (first entry)  
DE rAb Stage 1 VH construct.  
KW Humanised; human; mouse; recombinant; antibody; anti-VLA4; heavy;  
KW light; variable; framework; rAb; monoclonal; CDR; HPI/2; asthma; CDR;  
KW complementarity determining region; inflammatory bowel disease.  
OS Synthetic.  
FH Key  
FH peptide 1..19 Location/Qualifiers  
FT /note= "Signal peptide"  
FT protein 20..143  
FT /note= "Mature VH"  
FT region 27..29  
FT /note= "Murine residues"  
FT W09416094-A.  
PD 21-JUL-1994. U00266.  
PF 07-JAN-1994; US-004798.  
PR 12-JAN-1993; US-004798.  
PE (BIO.) BIOGEN INC.  
PE Carr FJ, Lobb RR, Tempest PR;  
PE WPI: 94-249240/30.  
PE N-PSDB; Q69901.  
PT Recombinant and humanised anti-VLA4 antibodies - contg.  
PT non-human CDRs in their light and heavy chains, used to treat  
PT (non-)specific inflammation, e.g. asthma  
PC Example 3: Page 90; 140pp; English.  
CC This sequence represents the stage 1 heavy chain variable region (VH)  
CC of the recombinant antibody (rAb) of the invention. The VH sequence has  
CC been CDR grafted and comprises a NEW framework and CDRs derived from  
CC the murine monoclonal antibody, HPI/2. In addition to CDR grafting, the  
CC stage 1 VH construct contains selected framework changes. Just prior to  
CC CDR1, a block of sequences have been changed to murine residues, Phe27,  
CC Asn28, Ile29 and Lys30. Although these four residues are not nominally  
CC included in CDR1, structurally they are included in part of the CDR loop,  
CC and are empirically included as part of CDR1. Arg was also changed to  
CC Asp at position 94. This sequence was used in the production of a  
CC humanised recombinant anti-VLA4 antibody (rAb). The humanised rAb  
CC produced comprises non-human CDRs at VH positions 31-35 (CDR1), 50-65





FH Key Location/Qualifiers  
 FT peptide 1..19  
 FT /note= "Signal peptide"  
 FT protein 20..143  
 FT /note= "Humanised HPI/2"  
 FT peptide 46..49  
 FT /note= "Murine residues"  
 FT peptide 50..54  
 FT /label= CDR1  
 FT peptide 59..85  
 FT /label= CDR2  
 FT misc\_difference 113  
 FT /note= "Murine residue"  
 FT peptide 118..129  
 FT /label= CDR3  
 PN WO9411027-A.  
 PD 26-MAY-1994.  
 PF 15-NOV-1993; U11060.  
 PR 13-NOV-1992; US-977702.  
 PA (UNIW ) UNIV WASHINGTON.  
 PI Papayannopoulou T;  
 DR WPI: 94-183163/22.  
 DR N-PSDB; Q65624.  
 PT Peripheralisation of CD34 positive stem cells in peripheral blood  
 PT - using blocking agent for VLA-4 surface antigen, useful for  
 PT reconstituting bone marrow after AIDS or cancer treatment and in  
 PT gene therapy  
 PS Disclosure: Page 36-37: 52pp; English.  
 CC The sequences given in R53208-09 represent the humanised variable  
 CC regions of the heavy and light chains of anti-VLA-4 murine monoclonal  
 CC antibody HPI/2. The Kabat NEMO framework was used to accept the heavy  
 CC chain CDRs and the Kabat REI framework was chosen to accept the kappa  
 CC chain CDRs. Transplantation of the CDRs was achieved using M13  
 CC mutagenesis vectors and synthetic oligonucleotides containing the  
 CC CDR-encoding sequences flanked by short sequences derived from the  
 CC frameworks. The VH sequence, in addition to the CDRs, contains murine  
 CC amino acids at positions 27-30 and an Arg to Asp change at position 94.  
 CC The VL sequence contains no further murine residues. The VLA-4 complex  
 CC is expressed at substantial levels on normal peripheral blood B and T  
 CC cells, thymocytes, monocytes and some melanoma cells as well as on  
 CC marrow blast cells and erythroblasts. The humanised HPI/2 antibody,  
 CC may be used in the peripheralisation of CD34 cells by acting as a  
 CC blocking agent of VLA-4 surface antigen. The isolated cells are  
 CC useful for autologous bone marrow reconstitution and as carriers in  
 CC gene therapy of acquired and inherited diseases. The peripheral-  
 CC isation method can also be used as a model for studying haematopoiesis,  
 CC homing of stem cells and cytokine-induced peripheralisation. Treatment  
 CC with the blocking agent increases the number of CD34 cells in peripheral  
 CC blood more effectively than treatment with cytokine.  
 SQ Sequence 143 AA;

Query Match 6.0%; Score 108; DB 10; Length 143;  
 Best Local Similarity 26.6%; Pred. No. 3.40e+00;  
 Matches 17; Conservative 19; Mismatches 24; Indels 4; Gaps 4;  
 Db 39 lctvsgfnikdtyhm-vwrgppg:glewigrdpasgdktydpkfqvrvtnlvdtsknq 97  
 QY 167 LDSTIDYFQPNKRNQLRLQTSN'DHVG-LGTAFENSKYDQDYNIRVTMYVQ-PR-E 223  
 Db 98 fsrl 101  
 QY 224 FNLK 227  
 RESULT 5  
 ID W72432 standard; Protein; 143 AA.  
 AC W72432;  
 DT 22-DEC-1998 (first entry)  
 DE Transplanted VH sequence.  
 KW peripheralising; CD34+ cell; haematopoietic stem cell; VLA-4 antigen;  
 KW cell proliferation; cancer; AIDS; gene therapy; chemotherapy;  
 KW acquired immune deficiency syndrome; radiotherapy; peripheral blood;  
 KW anti-VLA-4 antibody.

CS Synthetic.  
 JS Mus sr.  
 PN US5824304-A.  
 PD 20-OCT-1998.  
 PF 05-JUN-1995; 463298.  
 PR 15-NOV-1993; US-436339.  
 PR 13-NOV-1992; US-977702.  
 PR 05-JUN-1995; US-463298.  
 PA (PAPA) PAPAYANNOPOULOU T.  
 PI Papa/annopoulou T;  
 DI WPI: 98-582496/49.  
 DI N-FSDE; V66801.  
 DR Treatment of acquired immune deficiency syndrome by chemotherapy  
 DR and/or radiotherapy - with restoration of haematopoietic stem cells  
 DR enriched from peripheral blood after treatment with anti-VLA-4  
 DR antibody  
 PT Example 6: Column 19-20: 28pp; English.  
 PS A method has been developed for treating acquired immune deficiency  
 PS syndrome (AIDS). The method comprises: (1) peripheralising CD34+ cells  
 CC by administering an anti-VLA-4 antibody that blocks the binding of  
 CC VLA-4 antigen on the surface of the CD34+ cells to VCAM or fibronectin;  
 CC (2) collecting peripheral blood containing the CD34+ cells by  
 CC leucapheresis; (3) enriching the CD34+ cells by immunoadsorption using  
 CC anti-CD34 antibodies; (4) administering chemotherapy and/or radiotherapy  
 CC to the patient; and (5) returning the enriched CD34+ cells to the  
 CC patient's circulating blood. Experiments on macaques and baboons  
 CC indicate that the method increases peripheral blood leukocyte counts by  
 CC a factor of 2 and CFU-GM levels can be increased by a factor of 8-40.  
 CC The present sequence represents a transplanted VH sequence which is  
 CC used in an example from the present invention.  
 SQ Sequence 143 AA;

Query Match 6.0%; Score 108; DB 36; Length 143;  
 Best Local Similarity 26.6%; Pred. No. 3.40e+00;  
 Matches 17; Conservative 19; Mismatches 24; Indels 4; Gaps 4;  
 Db 39 lctvsgfnikdtyhm-vwrgppg:glewigrdpasgdktydpkfqvrvtnlvdtsknq 97  
 QY 167 LDSTIDYFQPNKRNQLRLQTSN'DHVG-LGTAFENSKYDQDYNIRVTMYVQ-PR-E 223  
 Db 98 fsrl 101  
 QY 224 FNLK 227  
 RESULT 6  
 ID W59645 standard; Protein; 325 AA.  
 AC W59645;  
 DT 12-OCT-1998 (first entry)  
 DE Amino acid sequence of the AS164.1 clone.  
 KW Clone AS164.1; secreted protein; human cell; immuno-modulator; PCR;  
 KW anti-tumour agent; tissue growth promoter; haemostatic agent; primer;  
 KW thrombolytic agent; molecular weight marker; genetic disorder;  
 KW antibody.  
 OS Homo sapiens.  
 PU WO9821332-A2.  
 PD 22-MAY-1998.  
 PF 14-NOV-1997; U20740.  
 PR 13-NOV-1997; US-969515.  
 PR 15-NOV-1996; US-749745.  
 PR 02-JUN-1997; US-367678.  
 PA (GENY ) GENETICS INST INC.  
 PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,  
 PI Racie LS, Spaulding V, Treacy M;  
 DR WPI: '98-297936/26.  
 DR N-PSDB; V41444.  
 PT New nucleic acid encoding secreted protein from human cells -  
 PT potentially useful, e.g. as immuno-modulators, anti-tumour agents,  
 PT promoters of tissue growth, haemostatic and thrombolytic agents  
 PT Claim 16; Pages 66-67; 98pp; English.  
 CC This is the amino acid sequence of the clone AS164-1, used in the  
 CC method of the invention involving secreted proteins from human cells  
 CC potentially useful as immuno-modulators, anti-tumour agents, tissue



PF 07-JAN-1994; U00266.  
PR 12-JAN-1993; US-004798.  
PA (BIOV) BIOGEN INC.

PI Carr FJ, Lobb RR, Tempest PR;  
DR WPI: 94-249240/30.  
DR N-PSDB: 069902.

PT Recombinant and humanised anti-VLA4 antibodies - contg.  
PT non-human CDRs in their light and heavy chains, used to treat  
PT (non-specific inflammation, e.g. asthma  
PS Example 4; Page 94; 140pp; English.

CC This sequence represents the stage 2 heavy chain variable region (VH)  
CC of the recombinant antibody (rAb) of the invention. This sequence is  
CC the same as the stage 1 construct (see also R59940) except for a  
CC substitution of a Ser for Lys at position 76. The VH sequence has  
CC been CDR grafted and comprises a NEW framework and CDRs derived from  
CC the murine monoclonal antibody, HPI/2. In addition to CDR grafting, the  
CC stage 1 VH construct contains selected framework changes. Just prior to  
CC CDR1, a block of sequences have been changed to murine residues, Phe27,  
CC Asn28, Ile29 and Lys30. Although these four residues are not nominally  
CC included in CDR1, structurally they are included in part of the CDR loop,  
CC and are empirically included as part of CDR1. Arg was also changed to  
CC Asp at position 94. This sequence was used in the production of a  
CC humanised recombinant anti-VLA4 antibody (rAb). The humanised rAb  
CC produced comprises non-human CDRs at VH positions 31-35 (CDR1), 50-65  
CC (CDR2) and 95-102 (CDR3). It may further comprise non-human residues  
CC at framework positions 27-30, 75, 77-79 or 66-67, or 69-71 and 69-71,  
CC or 84-85, or 38 and 40 or 24. The VL of the rAb has non-human CDRs at  
CC positions 24-34 (CDR1), 50-56 (CDR2) and 89-97 (CDR3). It may also have  
CC non-human residues at framework positions 60 and 67. The non-human CDRs  
CC are derived from HPI/2, which is a member of Kabat V with no unusual  
CC residues. The rAb can be used to treat inflammation in mammals, esp.  
CC asthma or inflammatory bowel disease.  
SQ Sequence 143 AA;

Query Match 5.7%; Score 103; DB 11; Length 143;

Best Local Similarity 27.3%; Pred. No. 7.85e+00;

Matches 15; Conservative 15; Mismatches 23; Indels 2; Gaps 2;

Db 39 lctcvgfnikdtyh-vwrgppgrglwgridpasgtdkydkfqrvtmldv 92

OY 167 LDSITIDYFPQNNKQLWLRLQTSQVNDVHG-LGTAFFNSKYDQDYNRYVMYQ 220

RESULT 10

ID R65199 standard; Protein: 449 AA.

AC R65199;

DT 19-AUG-1995 (first entry)

DE VPI protein.

KW Chicken anemia virus VPI protein; apoptosis; cancer therapy;

OS vaccine; antitumor; antibody generation.

PN Chicken anemia virus.

PN W09503414-A.

PD 02-FEB-1995.

PF 19-JUL-1994; NL0168.

PR 20-JUL-1993; NL-001272.

PA (AESC-) AESCULAAP BV.

PI Koch G, Noteborn NHH;

DR WPI: 95-075240/10.

DR N-PSDB: Q82828.

PT Chicken anaemia virus (CAV) mutant polypeptide(s) - useful as

PT vaccines or for inducing apoptosis

PS Claim 1; Fig 1; 53pp; English.

CC The sequence corresponds to a VPI protein from chicken anemia virus,

CC (CAV), and may be used to induce apoptosis directly or to generate

CC antibodies against CAV. The protein may be used as a vaccine or

CC an antitumor agent.

SQ Sequence 449 AA;

Query Match

Best Local Similarity 43.8%; Pred. No. 9.27e+00;

Matches 21; Conservative 9; Mismatches 13; Indels 5; Gaps 5;

Db 8 prgrfysfirg-whlklr-lrrrykfrhrrryrrafrkafnpr 53

OY 4 PARR-YRRRRHPRSHLGQILRRRPWLHPR-HRYWRR-KNGIFNTR 48

RESULT 11

ID R29580 standard; Protein: 657 AA.

AC R29580;

DT 22-APR-1993 (first entry)

DE FMR-1 gene product.

KW Fragile X disease; sex chromosome; X chromosome; X linked syndrome;

KW X linked retardation; X linked manic depression; TKCR;

KW Martin-Bell syndrome; CA polymorphisms; PCR analysis; ss.

OS Homo sapiens.

FH Key Location/Qualifiers

FT cds 1..3766

FT /\*tag= a

PN W09220825-A.

PN 26-NOV-1992.

PF 22-MAY-1992; U04447.

PR 24-MAY-1991; US-705490.

PR 29-AUG-1991; US-751891.

PA (BAYU) FAYLOR COLLEGE MEDICINE.

PA (UYEM-V) UNIV EMORY SCHOOL.

PA MEDICINE.

PI Caskey CJ, Nelson DL, Oostra BA, Pieretti M, Warren ST;

DR WPI: 92-415801/50.

DR N-PSDB: Q31890.

PT Gene sequence, related probes and cosmid(s) - useful in

PT diagnosing fragile X syndrome, X-linked mental retardation, manic

PT depression, and Martin Bell syndrome

PS Claim 14; Page 34; 75pp; English.

CC This sequence represents the FMR-1 gene product. It was isolated

CC as detailed in Q31890. The level of expression of the FMR-1 gene

CC may be determined by determining the level of this protein. This is

CC useful when trying to detect fragile X syndrome, X-linked retardation,

CC X-linked manic depression, TKCR and Martin-Bell syndrome.

SQ Sequence 657 AA;

Query Match

Best Local Similarity 5.7%; Score 103; DB 6; Length 657;

Matches 14; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

LL 18 PRRYRRRRHPRSHLGQILRRRPWL 43

OY 4 PRRYRRRRHPRSHLGQILRRRPWL 29

RESULT 12

ID W72434 standard; Protein: 143 AA.

AC W72434;

DT 22-DEC-1998 (first entry)

DE Humanised VH sequence.

KW peripheralising; CH34+ cell; haematopoietic stem cell; VLA-4 antigen;

KW cell proliferation; cancer; AIDS; gene therapy; chemotherapy;

KW acquired immune deficiency syndrome; radiotherapy; peripheral blood;

KW anti-VLA-4 antibody.

OS Synthetic.

OS Homo sapiens.

OS Mus. sp.

PN US5824304-A.

PN 20-OCT-1998.

PF 05-JUN-1995; 463298.

PR 15-NOV-1993; US-436339.

PR 13-NOV-1992; US-977702.

PR 05-JUN-1995; US-463298.

PA (PAPA/) PAPAYANNOPOULOU T.

PI Papayannopoulou T;

DR WPI: 98-582496/49.

DR N-PSDB: V66803.

PT Treatment of acquired immune deficiency syndrome by chemotherapy

PT and/or radiotherapy - with restoration of haematopoietic stem cells

PT enriched from peripheral blood after treatment with anti-VLA-4

PT antibody

Example 6; Column 23-26; 28pp; English.

A method has been developed for treating acquired immune deficiency syndrome (AIDS). The method comprises: (1) peripheralising CD34+ cells by administering an anti-VLA-4 antibody that blocks the binding of VLA-4 antigen on the surface of the CD34+ cells to VCAM or fibronectin; (2) collecting peripheral blood containing the CD34+ cells by leukapheresis; (3) enriching the CD34+ cells by immunoadsorption using anti-CD34 antibodies; (4) administering chemotherapy and/or radiotherapy to the patient; and (5) returning the enriched CD34+ cells to the patient's circulating blood. Experiments on macaques and baboons indicate that the method increases peripheral blood leukocyte counts by a factor of 2 and CFU-GM levels can be increased by a factor of 8-40. The present sequence represents a humanised VK sequence which is used in an example from the present invention.

Sequence 143 AA;

```

Query Match      5.68; Score 101; DB 36; Length 143;
Best Local Similarity 31.68; Pred. No. 1.09e-01;
Matches 12; Conservative 11; Mismatches 14; Indels 1; Gaps 1;

Db 55 wvrqpggrglewigrldpdsagdtkydpkfqrvtmlyd 92
      : : : : : : : : : : : : : : : : : : : :
QY 184 WLRLQTSNGVDHVG-LGTAFENSKYQDDYNIRVTMYVQ 220

```

RESULT	13	
ID	R55210	standard; Protein: 143 AA.
AC	R55210;	
DT	24-JAN-1995	(first entry)
DE	Humanised HP1/2	heavy chain variable region.
DW	Variable; region;	heavy; light; chains; anti-VLA-4; murine; monocytes;
KW	monoclonal antibody;	HP1/2; VLA-4 complex; peripheral blood; B cells;
KW	T cells; thymocytes;	Melanoma cells; marrow blast cells; haematopoiesis;
KW	erythroblasts; peripheralisation;	CD34 cells; blocking agent; cytokine;
KW	VLA-4 surface antigen;	bone marrow reconstitution; gene therapy;
KW	autologous; homing;	stem cells; cytokine-induced peripheralisation;
KW	Kabat; NEWM;	framework; REI; kappa; transplantation; M13 mutagenesis;
KW	KW	vectors.

OS	Chimeric - Mus musculus.
OS	Chimeric - Homo sapiens.
FS	Location/Qualifiers
FT	peptide
FT	1..19
FT	/note= "Signal peptide"
FT	protein
FT	20..143
FT	/note= "Humanised HP1/2"
FT	misc_difference 43
FT	/note= "Murine residue"
FT	peptide
FT	50..54
FT	/label= CDR1
FT	peptide
FT	69..85
FT	/label= CDR2
FT	misc_difference 94
FT	/note= "Murine residue"
FT	peptide
FT	118..129
FT	/label= CDR3
FT	FT

PN WO9411027-A.  
PD 26-MAY-1994.  
PF 15-NOV-1993. U11060.  
PR 13-NOV-1992; US-977702.  
PA (UNIW ) UNIV WASHINGTON.  
PR Papayannopoulou T;  
PI WPI; 94-183163/22.  
DR N-P8DB; Q65626.  
PT Peripherisation of CD34 positive stem cells in peripheral blood  
PT - using blocking agent for VLA-4 surface antigen, useful for  
PT re-constituting bone marrow after AIDS or cancer treatment and in  
PT gene therapy.  
PS Disclosure: Page 40-41: 52pp; English.  
CC The sequences given in R35210-11 represent the humanised variable  
CC regions of the heavy and light chains of anti-VLA-4 murine monoclonal  
CC antibody HPI/2. The Kabat NEMO framework was used to accept the heavy  
CC chain CDRs and the Kabat REI framework was chosen to accept the kappa  
CC chain CDRs. Transplantation of the CDRs was achieved using M13  
CC phage display.

mutagenesis vectors and synthetic oligonucleotides containing the CDR-encoding sequences flanked by short sequences derived from the frameworks. The VH sequence includes a Val to Ala change at position 24 and a Lys to Ser mutation at position 75 corresponding to murine amino acids. The VL sequence contains no further murine residues. The VLA-4 complex is expressed at substantial levels on normal peripheral blood B and T cells, thymocytes, monocytes and some melanoma cells as well as on marrow blast cells and erythroblasts. The humanised HPI/2 antibody, may be used in the peripheralisation of CD34 cells by acting as a blocking agent of VLA-4 surface antigen. The isolated cells are useful for autologous bone marrow reconstitution and as carriers in gene therapy of acquired and inherited diseases. The peripheralisation method can also be used as a model for studying haematopoiesis, homing of stem cells and cytokine-induced peripheralisation. Treatment with the blocking agent increases the number of CD34 cells in peripheral blood more effectively than treatment with cytokine.

```
Query Match      5.6%; Score 101; DB 10; Length 143;
Best Local Similarity 31.6%; Pred. No. 1.09e+01;
Matches 12; Conservative 11; Mismatches 14; Indels 1; Gaps 1;
```

Db 55 wvrqpqrqlwriqridpasqdtkydypkfqrvtmldv 92

QV 184 WLRLTSGNVDPHG-LGTAFENSKYDODYNIRVTMYVO 220

RESULT 14  
ID: W31690 standard; Protein: 143 AA.

AC	W31600 standard; 10000; 495 nm.
AD	W31690;
DE	14-APR-1998 (first entry)
DE	Humanised anti-VLA-4 MAb hsp1/2 variable region (Vh) sequence.
DE	anti-VLA-4; monoclonal antibody; MAb; human; CD34+ cell; treatment;
KW	cancer; VLA-4 antigen; chemotherapy.
KW	Homo sapiens.
OS	

OS	Key	Location/Qualifiers
FT	Peptide	1..19
FT	Protein	/note= "signal peptide" 20..143
FT		/note= "mature peptide"

FI	US5695755-A.	
PD	09-DEC-1997.	
PF	05-JUN-1995; 463128.	
PR	15-NOV-1993; US-4363309.	
PR	13-NOV-1992; US-977702.	
PR	05-JUN-1995; US-463128.	
PI	(PAPA/) PAPAYANNOPOULOU T.	
PI	Papayannopoulos T;	
DR	WPI; -98-041234/04.	
DR	N-PSPB: V02235.	

CC Cancer treatment involving extracorporeal separation of CD34+ cells  
 CC PT - after peripheralisation by administering anti-VLA-4 antibody  
 CC PS Example 6; Columns 23-24; 28pp; English.  
 CC CC This is the variable region of the heavy and light chains of a humanised  
 CC CC anti-VLA-4 monoclonal antibody hnp1/2. The anti-VLA-4 antibodies can be  
 CC CC used in a method for treating cancer in a patient. The method comprises  
 CC CC peripheralising CD34+ cells by administering the anti-VLA-4 antibody.  
 CC CC This antibody blocks the binding of VLA-4 antigen on the surface of the  
 CC CC CD34+ cells to VCAM or fibronectin. Peripheral blood containing the CD34+  
 CC CC cells is collected from a patient by leukapheresis. The CD34+ cells is  
 CC CC enriched by immunoadsorption using anti-CD34 antibodies. Chemotherapy  
 CC CC and/or radiotherapy can be administered to the patient and the enriched  
 CC CC CD34+ cells is returned to the patient's circulating blood. The method is  
 CC CC especially used for restoring haematopoietic stem cells to patients who  
 CC CC have undergone myeloablative chemotherapy and radiotherapy.  
 CC Sequence 143 AA;  
 SO

Query Match                    5.6%      Score 101; DB 27; Length 143;  
Best Local Similarity 31.6%      Pred. No. 1.09e+01;  
Matches                    12: Conservative      11: Mismatches 14: Indels

```
Db 55 wrqppgrglewigrldpasgdtkydkpfqvrvtmld 92
      |:| : : : : | :|| : :||| :
      |:| : : : : | :|| : :||| :
```

Search completed: Wed Dec 22 10:16:18 1999  
Job time : 23 secs.

\*\*\*\*\*  
WORLD  
\*\*\*\*\*  
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Wed Dec 22 10:17:49 1999; MasPar time 5.91 Seconds  
Tabular output not generated.

Title: >US-09-209-961-5  
Description: (1-233) from US09209961.pep  
Perfect Score: 1805  
Sequence: 1 MYPRRYRRRRRPRSHLG.....RVMTYVQFRENLKDPPLP 233

Scoring table: PAM 150  
Gap 11

Searched: 122461 seqs, 11912985 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-issued  
1:5A\_COMB 2:5B\_COMB 3:PCT9\_COMB 4:backfiles1

Statistics: Mean 31.451; Variance 147.510; scale 0.213

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	108	6.0	143	2	US-08-436- Sequence 4, Applicatio	2.26e+00
2	108	6.0	143	1	US-08-463- Sequence 4, Applicatio	2.26e+00
3	108	6.0	143	2	US-08-463- Sequence 4, Applicatio	2.26e+00
4	102	5.7	449	2	US-08-911- Sequence 3, Applicatio	5.96e+00
5	102	5.7	449	2	US-08-489- Sequence 3, Applicatio	5.96e+00
6	101	5.6	143	2	US-08-436- Sequence 8, Applicatio	7.00e+00
7	101	5.6	143	2	US-08-463- Sequence 8, Applicatio	7.00e+00
8	101	5.6	143	1	US-08-463- Sequence 8, Applicatio	7.00e+00
9	100	5.5	449	1	US-07-917- Sequence 2, Applicatio	8.21e+00
10	95	5.3	416	2	US-08-698- Sequence 5, Applicatio	1.81e+01
11	96	5.3	1087	1	US-08-264- Sequence 5, Applicatio	1.55e+01
12	93	5.2	124	2	US-08-657- Sequence 16, Applicati	2.47e+01
13	94	5.2	183	4	5168049-4 Patent No. 5168049.	2.11e+01
14	93	5.2	316	1	US-08-656- Sequence 1, Applicatio	2.47e+01
15	93	5.2	316	1	US-08-038- Sequence 1, Applicatio	2.47e+01
16	93	5.2	1257	1	US-08-340- Sequence 49, Applicati	2.47e+01
17	90	5.0	77	2	US-08-726- Sequence 174, Applicat	3.93e+01
18	90	5.0	432	2	US-08-933- Sequence 47, Applicati	3.93e+01
19	91	5.0	614	3	PCT-US95-0 Sequence 21, Applicati	3.37e+01
20	91	5.0	2291	2	US-08-286- Sequence 29, Applicati	3.37e+01
21	88	4.9	124	2	US-08-657- Sequence 17, Applicati	5.34e+01
22	89	4.9	559	2	US-08-406- Sequence 20, Applicati	4.58e+01
23	89	4.9	1754	2	US-08-311- Sequence 13, Applicati	4.58e+01

24	89	4.9	1754	1	US-07-745- Sequence 13, Applicati	4.58e+01
25	86	4.8	116	3	PCT-US95-0 Sequence 41, Applicati	7.23e+01
26	86	4.8	116	2	US-08-561- Sequence 41, Applicati	7.23e+01
27	87	4.8	119	1	US-08-474- Sequence 65, Applicati	6.22e+01
28	87	4.8	119	1	US-08-474- Sequence 89, Applicati	6.22e+01
29	87	4.8	119	1	US-08-477- Sequence 89, Applicati	6.22e+01
30	87	4.8	119	1	US-07-634- Sequence 89, Applicati	6.22e+01
31	87	4.8	119	1	US-08-477- Sequence 65, Applicati	6.22e+01
32	87	4.8	119	1	US-08-487- Sequence 65, Applicati	6.22e+01
33	87	4.8	119	1	US-07-634- Sequence 65, Applicati	6.22e+01
34	87	4.8	119	1	US-08-487- Sequence 89, Applicati	6.22e+01
35	86	4.8	135	2	US-08-137- Sequence 100, Applicat	7.23e+01
36	86	4.8	135	2	US-08-436- Sequence 103, Applicat	7.23e+01
37	87	4.8	237	2	US-08-910- Sequence 13, Applicati	6.22e+01
38	87	4.8	244	2	US-08-910- Sequence 22, Applicati	6.22e+01
39	84	4.7	135	2	US-08-436- Sequence 102, Applicat	9.77e+01
40	84	4.7	135	2	US-08-137- Sequence 102, Applicat	9.77e+01
41	85	4.7	380	2	US-08-902- Sequence 3, Applicatio	8.41e+01
42	85	4.7	1040	2	US-08-254- Sequence 2, Applicatio	8.41e+01
43	83	4.6	139	2	US-08-253- Sequence 19, Applicati	1.14e+02
44	83	4.6	1319	2	US-08-290- Sequence 2, Applicatio	1.14e+02
45	83	4.6	1708	2	US-08-508- Sequence 2, Applicatio	1.14e+02

ALIGNMENTS

RESULT 1  
ID US-08-436-339A-4 STANDARD; PRT; 143 AA.

XX xxxxxx

Sequence 4, Application US/08436339A

Sequence 4, Application US/08436339A

Patent No. 5843438

GENERAL INFORMATION:  
APPLICANT: Papayannopoulou, Thalia (USA only)

APPLICANT: Board of Regents, U.

TITLE OF INVENTION: PERIPHERALIZATION OF HEMATOPOIETIC STEM

TITLE OF INVENTION: CELLS

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: C/O FISH & NEAVE

STREET: 1251 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10020

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC Compatible

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/436.339A

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/463,128

FILING DATE: 05-JUN-1995

APPLICATION NUMBER: PCT/US93/11060

FILING DATE: 11-11-1993

APPLICATION NUMBER: US 07/977,702

FILING DATE: 13-NOV-1992

ATTORNEY/AGENT INFORMATION:

NAME: Haley Jr., James F.

REGISTRATION NUMBER: 27,794

REFERENCE/DOCKET NUMBER: B173CIP

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 596-9000





Db 39 LTCVSGFNKIDTYMH-VWRPPGRLGWIGRIDPASGDTKYDPKQVQVMTMLVDTSKNQ 97  
167 LDSTIDYFQPNKRQNLWLRLQTSQNVHDVG-LGTAENS KYDQDYNIRVTWVQ-FR-E 223  
Db 98 FSLR 101  
Qy 224 FNLK 227

RESULT 4  
ID US-08-911-092-3 STANDARD; PRT; 449 AA.

XXXXXX

Sequence 3, Application US/08911092

Sequence 3, Application US/08911092

Patent No. 5952002

GENERAL INFORMATION:

APPLICANT: No. 5952002eborn, Mathews H.M.

APPLICANT: Koch, Guus

TITLE OF INVENTION: Chicken Anemia Virus Mutants And Vaccines

TITLE OF INVENTION: And Uses Based On The Viral Proteins VP1, VP2, And VP3 Or

TITLE OF INVENTION: Sequences Of That Virus Coding Therefor

NUMBER OF SEQUENCES: 32

CORRESPONDENCE ADDRESS:

ADDRESSEE: Rae-Venter Law Group, P.C.

STREET: P.O. Box 60039

CITY: Palo Alto

STATE: California

COUNTRY: USA

ZIP: 94306

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/911,092

FILING DATE: 14-AUG-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/454,121

FILING DATE: 30-NOV-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/NL94/00168

FILING DATE: 19-JULY-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: NL 9301272

FILING DATE: 20-JULY-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/030,335

FILING DATE: 8-MAR-1993

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/NL91/00165

FILING DATE: 11-SEP-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: NL 9002008

FILING DATE: 12-SEP-1990

ATTORNEY/AGENT INFORMATION:

NAME: Rae-Venter, Barbara

REGISTRATION NUMBER: 32,750

REFERENCE/DOCKET NUMBER: LEBV003.00US1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (650) 328-4400

TELEFAX: (650) 328-4477

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

CC LENGTH: 449 amino acids

CC TYPE: amino acid

CC TOPOLOGY: unknown

CC MOLECULE TYPE: protein

CC HYPOTHETICAL: NO

CC SEQUENCE 449 AA; 51638 MW; 1077284 CN;

Query Match

Best Local Similarity 5.7%; Score 102; DB 2; Length 449;

Matches 21; Conservative 9; Mismatches 13; Indels 5; Gaps 5;

Db 8 PRGRYFSEHGR-WHLKR-LRRYKFRHRRORYRRRAFRKAFHNP 53

Qy 4 PRRR-YERRHRPRSHLQILRRRWLVHPR-HRYRWRR-KNGIFNTR 48

RESULT 5

ID US-08-489-666C-3 STANDARD; PRT; 449 AA.

XX AC XXXXXX

XX DT

XX XX

DE Sequence 3, Application US/08489666C

XX Sequence 3, Application US/08489666C

CC Patent No. 5922600

CC GENERAL INFORMATION:

CC APPLICANT: No. 5922600eborn, M.H.M

CC APPLICANT: Koch, G.

CC TITLE OF INVENTION: Chicken Anemia Virus mutants and

CC TITLE OF INVENTION: vaccines and uses based on the viral proteins VP1, VP2

CC TITLE OF INVENTION: VF3 or sequences of that virus coding therefor.

CC NUMBER OF SEQUENCES: 30

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: RAE-VENTER LAW, P.C.

CC STREET: 260 SHERIDAN AVE., P.O. BOX 60039

CC CITY: PALO ALTO

CC STATE: CA

CC COUNTRY: USA

CC ZIP: 94306

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patent In Release #1.0, Version #1.30

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/489,666C

CC FILING DATE: 07-JUN-1995

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: US 08/454,121

CC FILING DATE: 30-NOV-1995

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: US 08/030,335

CC FILING DATE: 08-MAR-1993

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: PCT/NL 94/00168

CC FILING DATE: 19-JUL-1994

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: PCT/NL 91/00165

CC FILING DATE: 11-SEP-1991

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: NL 9301272

CC FILING DATE: 20-JUL-1993

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: NL 9002008

CC FILING DATE: 12-SEP-1990

CC ATTORNEY/AGENT INFORMATION:

CC NAME: RAE-VENTER, BARBARA

CC REGISTRATION NUMBER: 32,750

CC REFERENCE/DOCKET NUMBER: LEBV.003.04US

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (650) 328-4400



ADDRESSEE: Incyte Pharmaceut  
STREET: 3174 Porter Drive

..



CC NAME: Halluin, Albert P.  
CC REGISTRATION NUMBER: 25,227  
CC REFERENCE/DOCKET NUMBER: 7606-028  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 415-854-3660  
CC TELEFAX: 415-854-3694  
CC TELEX: 66141 PENNIE  
CC INFORMATION FOR SEQ ID NO: 16:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 124 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: unknown  
CC TOPOLOGY: unknown  
CC MOLECULE TYPE: protein  
CC SEQUENCE 124 AA; 13972 MW; 91266 CN;  
Query Match 5.2%; Score 93; DB 2; Length 124;  
Best Local Similarity 26.3%; Pred. No. 2.47e+01;  
Matches 10; Conservative 14; Mismatches 13; Indels 1; Gaps 1;  
Db 36 WVRQPPGRGLEWIGVINYDDDTYNQKFKGRVTMLVD 73  
Qy 184 WLRLTSGNVHDVGLTAFEN-SKIDQYINIRVTWVQ 220  
RESULT 13  
ID 5168049-4 STANDARD; PRT; 198 AA.  
XX XXXXXX  
AC XXXXXX  
DT 01-JAN-1900  
XX  
DE Patent No. 5168049.  
XX  
XX Patent No. 5168049  
CC APPLICANT: MEADE, HARRY M.; GARWIN, JEFFREY L.  
CC TITLE OF INVENTION: PRODUCTION OF STREPTAVIDIN-LIKE  
CC POLYPEPTIDES  
CC NUMBER OF SEQUENCES: 6  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/07/185,329  
CC FILING DATE: 21-APR-1988  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 656,873  
CC FILING DATE: 02-OCT-1984  
CC SEQ ID NO.4:  
CC LENGTH: 183  
CC SEQUENCE 198 AA; 23791 MW; 163858 CN;  
Query Match 5.2%; Score 94; DB 4; Length 183;  
Best Local Similarity 41.0%; Pred. No. 2.11e+01;  
Matches 16; Conservative 7; Mismatches 14; Indels 2; Gaps 1;  
Db 86 RPHRRQRHPRDLGGLLEXPQLRDLRHPVRRRR 124  
Qy 5 RRRYRRRRRPR--SHLGQILRRRPLVHPRHRYRWRK 41  
RESULT 14  
ID US-08-656-349-1 STANDARD; PRT; 316 AA.  
XX XXXXXX  
AC XXXXXX  
XX  
DT  
DE  
XX  
DE Sequence 1, Application US/08656349  
CC Sequence 1, Application US/08656349  
CC Patent No. 5728544  
CC GENERAL INFORMATION:  
CC APPLICANT:  
CC TITLE OF INVENTION: NOVEL PROTEASE II  
CC NUMBER OF SEQUENCES: 7

CC -COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/656,349  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC INFORMATION FOR SEQ ID NO: 1:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 316 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 316 AA; 34331 MW; 527254 CN;  
Query Match 5.2%; Score 93; DB 1; Length 316;  
Best Local Similarity 27.8%; Pred. No. 2.47e+01;  
Matches 15; Conservative 15; Mismatches 21; Indels 3; Gaps 3;  
Db 33 NTRGNGIFTYDAKYRTTLPGSLWADADNOF-EASYDAPAVDAHYAGVTYDYK 85  
Qy 46 NTRLSTFGTVK-RTVTTPSWAVDMRMFKIDDFVPPGGTNN-ISIPFEYIR 97  
RESULT 15  
ID US-08-038-932B-1 STANDARD; PRT; 316 AA.  
XX XXXXXX  
AC XXXXXX  
DT  
XX  
DE  
XX  
DE Sequence 1, Application US/08038932B  
CC Sequence 1, Application US/08038932B  
CC Patent No. 5496710  
CC GENERAL INFORMATION:  
CC APPLICANT: Nagao, Hiromasa  
CC APPLICANT: Yoneya, Takashi  
CC APPLICANT: Miyake, Toshio  
CC APPLICANT: Aoyama, Atsuo  
CC APPLICANT: Rai, Ken-ichi  
CC APPLICANT: Kidokoro, Shun-ichi  
CC APPLICANT: Miki, Yoichiro  
CC APPLICANT: Endo, Kimiko  
CC APPLICANT: Wada, Akiyoshi  
CC TITLE OF INVENTION: NOVEL PROTEASE  
CC NUMBER OF SEQUENCES: 7  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Cushman, Darby & Cushman  
CC STREET: 1100 New York Avenue,  
CC CITY: N.W.  
CC STATE: Washington, D.C.  
CC COUNTRY: U.S.A.  
CC ZIP: 20005-3918  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/038,932B  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: JP 4-171479  
CC FILING DATE: 08-JUN-1992  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: JP 4-237606  
CC FILING DATE: 14-AUG-1992  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: JP 4-273413

CC FILING DATE: 18-SEP-1992

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: 202-861-3000

CC TELEFAX: 202-638-4203

CC TELEX: 248453

CC INFORMATION FOR SEQ ID NO: 1:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 316 amino acids

CC TYPE: amino acid

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

SQ SEQUENCE 316 AA; 34331 MW; 527254 CN;

Query Match

Best Local Similarity 5.28; Score 93; DB 1; Length 316;

Matches 15; Conservative 15; Mismatches 21; Indels 3; Gaps 3;

Db 33 NTRNGIFTYDAKVRTLPGLWADADNQF-EASYDAPAVDAHYAGVTVDYK 85

QY 46 NTRLSRTFGTVK-RTVTTPSWAVDMRKFIDFVPPGGTNK-ISISPEYR 97

Search completed: Wed Dec 22 10:17:59 1999  
Job time : 10 secs.

\*\*\*\*\*  
WATERMAN  
\*\*\*\*\*  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Mpsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Dec 22 10:16:35 1999; Maspar time 13.32 Seconds  
700.946 Million cell updates/sec

Tabular output not generated.

Title: >US-09-209-961-5  
Description: (1-233) from US09209961.pep  
Perfect Score: 1805  
Sequence: 1 MTPYRRRRRRRRSHLG.....RVTMVQFRFNKDPDPLEP 233

Scoring table: PAM 150  
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: pir60  
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 45.674; Variance 94.628; scale 0.483

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES							
Result No.	Score	Query Match	Length	ID	Description	Pred. No.	
1	158	8.8	8.8	118	2 S56117	spermatid-specific pr	8.95e-09
2	153	8.5	77	2 B40973	spermatid-specific pr	5.14e-08	
3	146	8.1	78	2 A40973	spermatid-specific pr	5.73e-07	
4	146	8.1	79	2 S56116	spermatid-specific pr	5.73e-07	
5	126	7.0	58	2 S34045	protamine - North Ame	4.29e-04	
6	120	6.6	1538	2 H70846	hypothetical glycine-	2.85e-03	
7	113	6.3	52	2 PNO081	sperm chromatin prote	2.44e-02	
8	113	6.3	57	2 A34356	protamine - Japanese	2.44e-02	
9	113	6.3	61	2 S39425	protamine p1 - duckbi	2.44e-02	
10	114	6.3	113	2 S65936	probable membrane pro	1.80e-02	
11	110	6.1	126	2 S58321	probable membrane pro	5.99e-02	
12	110	6.1	129	2 C69488	LSU ribosomal protein	5.99e-02	
13	110	6.1	1460	1 EDBE1F	immediate-early prote	5.99e-02	
14	108	6.0	87	2 S00180	spermatid protein S1	1.08e-01	
15	109	6.0	225	2 C71159	hypothetical protein	8.05e-02	
16	107	5.9	238	2 A57198	splicing factor, argi	1.45e-01	
17	106	5.9	284	2 S26437	hypothetical protein	1.94e-01	
18	105	5.8	31	1 CLHRZA	protamine 2 - Atlanti	2.59e-01	
19	105	5.8	31	1 CLHRZ	protamine II - Pacific	2.59e-01	
20	105	5.8	56	2 C58213	protamine - chicken	2.59e-01	
21	105	5.8	65	1 GACH	probable membrane pro	3.45e-01	
22	104	5.8	815	2 S67675	protamine St2b - hors	6.10e-01	
23	102	5.7	58	2 S10755			

24	102	5.7	62	2	B34326	protamine - chicken	6.10e-01
25	102	5.7	62	2	S10754	protamine St2a - hors	6.10e-01
26	102	5.7	252	2	C43679	ORF3 protein - Autogr	4.59e-01
27	102	5.7	449	2	C39926	hypothetical protein	6.10e-01
28	102	5.7	1791	2	102345	hypothetical protein	6.10e-01
29	101	5.6	102	2	S53336	protamine P2 - rhesus	8.08e-01
30	101	5.6	104	2	S53118	protamine P2 - common	8.08e-01
31	101	5.6	710	2	S54152	sepB protein - Emeric	8.08e-01
32	101	5.6	968	2	T00322	hypothetical protein	8.08e-01
33	100	5.5	62	2	S58213	protamine I - America	1.07e+00
34	100	5.5	163	2	I38191	nucleic acid binding	1.07e+00
35	100	5.5	499	2	J01175	hypothetical 51.8K pr	1.07e+00
36	99	5.5	474	2	S73677	hypothetical protein	1.41e+00
37	99	5.5	493	2	S61171	ferredoxin--NADP+ red	1.41e+00
38	100	5.5	828	2	S34695	hypothetical protein	1.07e+00
39	100	5.5	5232	2	A45086	HC-toxin synthetase -	1.07e+00
40	97	5.4	50	1	HSSH	sperm histone - sheep	2.45e+00
41	96	5.4	249	2	F71342	probable UDP-N-acetyl	1.86e+00
42	96	5.4	403	1	VCBBBL	coat protein precursor	1.86e+00
43	98	5.4	514	1	W2WLB5	E2 protein - human pa	1.86e+00
44	98	5.4	2424	2	I46479	calcium channel BI-2	1.86e+00
45	98	5.4	2424	2	I46480	calcium channel BI-2	1.86e+00

ALIGNMENTS

RESULT 1

ENTRY S56117 #type complete

TITLE Spermatid-specific protein T2 precursor - longfin squid

ALTERNATE\_NAMES sperm protamin Sp

ORGANISM #formal\_name Loligo pealeii #common\_name longfin squid

DATE 27-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 03-Nov-1995

ACCESSIONS S56117

REFERENCE S56116

#authors Wouters-Tyrou, D.; Martin-Ponthieu, A.; Ledoux-Andula, N.; Kouach, M.; Jaquinod, M.; Subirana, J.A.; Sautiere, P.

#journal Biochem. J. (1995) 309:529-534

#title Squid spermiogenesis: molecular characterization of testis-specific pro-protamines.

#accession S56117

#status preliminary

#molecule\_type protein

#residues 1-118 #label WOU

SUMMARY #length 118 #molecular-weight 16505 #checksum 2360

Query Match 8.8%; Score 158; DB 2; Length 118;  
Best Local Similarity 52.5%; Pred. No. 8.95e-09;  
Matches 21; Conservative 8; Mismatches 9; Indels 2; Gaps 2;

Db 68 SYSRRRRRRR-RSRSPYRRRRRRRRRRRRRRRRRRRRRRR 106  
QY :||||| ||||| ||||| : ||| : ||||| |||||  
2 TTPRRR-YRRRRRRRSHLGQILRRPWLVPVRRHRRWR 40

RESULT 2

ENTRY B40973 #type complete

TITLE spermatid-specific protein T2 precursor - common cuttlefish

ALTERNATE\_NAMES arginine-rich protamine; testis-specific protein T2

ORGANISM #formal\_name Sepia officinalis #common\_name common cuttlefish

DATE 21-Apr-1992 #sequence\_revision 21-Apr-1992 #text\_change 13-Sep-1998

ACCESSIONS B40973; S14086

REFERENCE B40973

#authors Wouters-Tyrou, D.; Chartier-Harlin, M.C.; Martin-Ponthieu, A.; Boutillon, C.; Van Dorsselaer, A.; Sautiere, P.

#journal J. Biol. Chem. (1991) 266:17388-17395

#title Cuttlefish spermatid-specific protein T. Molecular characterization of two variants T1 and T2, putative precursors of sperm protamine variants Sp1 and Sp2.

#cross-references MUID:91373359

#accession B40973

#molecule\_type protein

```
##residues 1-77 ##label WOU
REFERENCE S14085
#authors Martin-Ponthieu, D.; Wouters-Tyrou, D.; Belaiche, D.;
#journal Sautiere, P.; Schindler, P.; van Dorsselaer, A.
#title Eur. J. Biochem. (1991) 195:611-619
#cross-references MUID:91153298
#accession S14086
##residues 22-77 ##label MAR
CLASSIFICATION #superfamily sperm histone
KEYWORDS DNA binding; nucleus; phosphoprotein; spermatogenesis
FEATURE 1-21
22-77 #domain signal sequence #status predicted #label SIG\
#product protamine variant Sp2 #status experimental
#label MAT
SUMMARY #length 77 #molecular-weight 10485 #checksum 2088
Query Match 8.5%; Score 153; DB 2; Length 77;
Best Local Similarity 55.3%; Pred. No. 5.14e-08;
Matches 21; Conservative 7; Mismatches 7; Indels 3; Gaps 3;
Db 43 Y-RRRYRRRRRRRRRRR-R-SYRRRYRRRR 77
QY 3 YPRRRRRRRRSHLGQILRRPWLVPVPHRYRWRR 40
RESULT 3
ENTRY A40973 #type complete
TITLE spermatid-specific protein T1 precursor - common cuttlefish
ALTERNATE_NAMES arginine-rich protamine; testis-specific protein T1
ORGANISM #formal_name Sepia officinalis #common_name common cuttlefish
DATE 21-Apr-1992 #sequence_revision 21-Apr-1992 #text_change
13-Sep-1998
ACCESSIONS A40973; S14085
REFERENCE Wouters-Tyrou, D.; Chartier-Harlin, M.C.; Martin-Ponthieu,
#authors A.; Boutillon, C.; Van Dorsselaer, A.; Sautiere, P.
#journal J. Biol. Chem. (1991) 266:17388-17395
#title Cuttlefish spermatid-specific protein T1. Molecular
characterization of two variants T1 and T2, putative
precursors of sperm protamine variants Sp1 and Sp2.
#cross-references MUID:91373359
#accession A40973
#molecule_type protein
##residues 1-78 ##label WOU
REFERENCE S14085
#authors Martin-Ponthieu, D.; Wouters-Tyrou, D.; Belaiche, D.;
#journal Sautiere, P.; Schindler, P.; van Dorsselaer, A.
#title Eur. J. Biochem. (1991) 195:611-619
#cross-references MUID:91153298
#accession S14086
##residues 22-78 ##label MAR
CLASSIFICATION #superfamily sperm histone
KEYWORDS DNA binding; nucleus; phosphoprotein; spermatogenesis
FEATURE 1-21
22-78 #domain signal sequence #status predicted #label SIG\
#product protamine variant Sp1 #status experimental
#label MAT
SUMMARY #length 78 #molecular-weight 10631 #checksum 3554
Query Match 8.1%; Score 146; DB 2; Length 78;
Best Local Similarity 50.0%; Pred. No. 5.73e-07;
Matches 19; Conservative 8; Mismatches 9; Indels 2; Gaps 2;
Db 43 Y-RRRYRRRRRRRRRRR-SRRRYRRRRSRRRYRRRR 78
QY 3 YPRRRRRRRRSHLGQILRRPWLVPVPHRYRWRR 40
```

```
RESULT 4
ENTRY S56116 #type complete
TITLE spermatid-specific protein T1 - longfin squid
ORGANISM #formal_name Loligo pealeii #common_name longfin squid
DATE 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change
03-May-1996
ACCESSIONS S56116
REFERENCE Wouters-Tyrou, D.; Martin-Ponthieu, A.; Ledoux-Andula, N.;
#authors Kouach, M.; Jaquinod, M.; Subirana, J.A.; Sautiere, P.
#journal Biochem. J. (1995) 309:529-534
#title Squid spermiogenesis: molecular characterization of
testis-specific pro-protamines.
#accession S56116
##status preliminary
##molecule_type protein
##residues 1-79 ##label WOU
CLASSIFICATION #superfamily sperm histone
SUMMARY #length 79 #molecular-weight 10788 #checksum 5396
Query Match 8.1%; Score 146; DB 2; Length 79;
Best Local Similarity 50.0%; Pred. No. 5.73e-07;
Matches 19; Conservative 8; Mismatches 9; Indels 2; Gaps 2;
Db 44 Y-RRRYRRRRRRRRR-SRRRYRRRRSRRRYRRRRR 79
QY 3 YPRRRRRRRRSHLGQILRRPWLVPVPHRYRWRR 40
RESULT 5
ENTRY S34045 #type complete
TITLE protamine - North American opossum
ORGANISM #formal_name Didelphis virginiana, Didelphis marsupialis
DATE 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change
08-Sep-1997
ACCESSIONS S34045
REFERENCE Winkfein, R.J.; Nishikawa, S.; Connor, W.; Dixon, G.H.
#authors Eur. J. Biochem. (1993) 215:63-72
#journal Characterization of a marsupial sperm protamine gene and its
transcripts from the North American opossum (Didelphis
marsupialis).
#cross-references MUID:93345500
#accession S34045
##status preliminary
##molecule_type DNA
##residues 1-58 ##label WIN
CLASSIFICATION #superfamily sperm histone
KEYWORDS DNA binding; nucleus
SUMMARY #length 58 #molecular-weight 7941 #checksum 3630
Query Match 7.0%; Score 126; DB 2; Length 58;
Best Local Similarity 50.0%; Pred. No. 4.29e-04;
Matches 19; Conservative 8; Mismatches 8; Indels 3; Gaps 3;
Db 3 RYRRRSRSRSRYGRR-RRRRSRRRRSR-RRRRRRG 38
QY 7 RYRR-RRRRSHLGQILRRPWLVPVPHRYRWRRKNG 43
RESULT 6
ENTRY H70846 #type complete
TITLE hypothetical glycine-rich protein RV3345c - Mycobacterium
tuberculosis (strain H37Rv)
ORGANISM #formal_name Mycobacterium tuberculosis
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
24-Jul-1998
ACCESSIONS H70846
REFERENCE Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
```



```

08-Sep-1997
ACCESSION   A34356
REFERENCE   A34356
#authors    3
#journal    J. Biol. Chem. (1989) 264:17627-17630
#title      Quail (Coturnix japonica) protamine, full-length cDNA
            sequence, and the function and evolution of vertebrate
            protamines.
#cross-references MUID:90036816
#accession  A34356
#status     preliminary
#molecule_type mRNA
#residues   1-57 #label OLI
##cross-references GB:M30275; NID:g213612; PID:g213613
CLASSIFICATION #superfamily sperm histone
KEYWORDS      DNA binding; nucleus; phosphoprotein
SUMMARY       #length 57 #molecular-weight 7828 #checksum 5804

Query Match          6.3%; Score 113; DB 2; Length 57;
Best Local Similarity 46.2%; Pred. No. 2,44e-02;
Matches 18; Conservative 7; Mismatches 11; Indels 3; Gaps 3;

Db 16 RRSRRRSRRRRYGRS-RRSVRSVGRRRRRYGRRRR 53
||| ||||| : : : : : : : : : : : : : : : : : : :
CY 5 RRPYRRRRRPSHLGQLRRPWLV-HPRHRY-RWRRK 41

RESULT 9
ENTRY   S39425 #type complete
TITLE   protamine PI - duckbill platypus
ORGANISM #formal_name Canithorhynchus anatinus #common_name duckbill
            platypus
DATE    29-Jan-1999
#cross-references EMBL:Z26849

ACCESSIONS S39425
REFERENCE   S39424
#authors    Retief, J.D.; Winkfein, R.J.; Dixon, G.H.
#journal    Eur. J. Biochem. (1993) 218:457-461
#title      Evolution of the monotremes. The sequences of the protamine
            PI genes of platypus and echidna.
#cross-references MUID:94094837
#accession  S39425
#status     preliminary
#molecule_type DNA
#residues   1-61 #label RET
##cross-references EMBL:Z26849

GENETICS    47/1
#introns
CLASSIFICA"ICN #superfamily sperm histone
SUMMARY     #length 61 #molecular-weight 7871 #checksum 4490

Query Match          6.3%; Score 113; DB 2; Length 61;
Best Local Similarity 43.8%; Pred. No. 2,44e-02;
Matches 21; Conservative 10; Mismatches 14; Indels 3; Gaps 3;

Db 11 SRSLYRRRR-RRR-RGGRTSRK-LSRSGRRRRRRRSGRRSSR 55
: ||||| : : : : : : : : : : : : : : : : : :
CY 4 PRRYRRRRRPSHLGQLRRPWLV-HPRHRYRWKKGIENTLSR 51

RESULT 10
ENTRY   S66936 #type complete
TITLE   probable membrane protein YOR053w - yeast (Saccharomyces
            cerevisiae)
ALTERNATE_NAMES hypothetical protein O2799
ORGANISM #formal_name Saccharomyces cerevisiae
DATE     12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change
            21-Nov-1997
ACCESSIONS S66936; S66927
REFERENCE   S66929
#authors    Bohn, C.; Bolotin-Fukuhara, M.; Daigman-Fornier, B.; Dang,
            D.V.; Valens, M.
#submission submitted to the Protein Sequence Database, July 1996

```

```

KEYWORDS          transmembrane protein
FEATURE
  3-19             #domain transmembrane #status predicted #label TM1\
  107-123          #domain transmembrane #status predicted #label TM2
SUMMARY           length 126 #molecular-weight 16294 #checksum 1384

Query Match      6.1%; Score 110; DB 2; Length 126;
Best Local Similarity 31.7%; Pred. No. 5.99e-02;
Matches 13; Conservative 14; Mismatches 14; Indels 0; Gaps 0;

Db    49 KRTRKRRKRKKRKKRKKRKKRKKRKKRKKRKNDAFY 89
QY    5 RRRYRRHRPRLSHLGQILRRRPWLVPVHPHRYRWRRKNGIF 45
      :| | | | : | : | : | | | | : | : | :
      || ||||| | | | | | | | | | | | | | | |

RESULT 12
ENTRY   C69488      #type complete
TITLE   LSU ribosomal protein L32E (rpl32B) homolog - Archaeoglobus
        fulgidus
ORGANISM #formal_name Archaeoglobus fulgidus
DATE     05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
        05-Jun-1998
ACCESSIONS C69488
REFERENCE A69250
AUTHORS    Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson,
            K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.;
            Peterson, J.D.; Richardson, D.L.; Kerlavage, A.R.; Graham,
            D.E.; Kyripides, N.C.; Fleischmann, R.D.; Quackenbush, J.;
            Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;
            Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.;
            Peterson, S.; Reich, C.I.; McNeill, L.K.; Badger, J.H.;
            Glodex, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman,
            J.F.; McDonald, L.; Utterback, T.; Cotton, M.D.; Spriggs,
            T.; Artiaich, P.; Kaine, B.P.; Sykes, S.M.; Sadov, P.W.;
            D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.;
            Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese,
            C.R.; Venter, J.C.
#journal Nature (1997) 390:364-370
#title The complete genome sequence of the hyperthermophilic,
        sulfate-reducing archaeon Archaeoglobus fulgidus.
#cross-references MWID:98049343
#accession C69488
#status preliminary; nucleic acid sequence not shown;
        translation not shown
#molecule_type DNA
#residues 1-129 #label KLE
#cross-references GB:AE000971; GB:AE000782; NID:g2689294; PID:g2648647;
        TIGR:AF1908
CLASSIFICATION #superfamily rat ribosomal protein L32
SUMMARY        #length 129 #molecular-weight 15151 #checksum 3045

Query Match      6.1%; Score 110; DB 2; Length 129;
Best Local Similarity 31.4%; Pred. No. 5.99e-02;
Matches 16; Conservative 14; Mismatches 18; Indels 3; Gaps 3;

Db    3 HKRLKVRQRQARKPEF-RRYCNWKKLRNKSRRPRGLFN-KLKRRYG 51
QY    5 RRRYRRHRPRLSHLGQILRRRPWLVPVHPHRYRWRRKNGIFRLSRTG 54
      :| | | | : | : | : | | | | : | | | : |
      || ||||| | | | | | | | | | | | | | | |

RESULT 13
ENTRY   EDBeIF      #type complete
TITLE   immediate-early protein IE180 - suid herpesvirus 1 (strain
        Indiana-Funkhauser)
ORGANISM #formal_name suid herpesvirus 1
DATE     30-Jun-1990 #sequence_revision 30-Jun-1990 #text_change
        16-Feb-1997
ACCESSIONS S04713
REFERENCE S04713
AUTHORS    Cheung, A.K.
#journal Nucleic Acids Res. (1989) 17:4637-4646
#title DNA nucleotide sequence analysis of the immediate-early gene
        of pseudorabies virus.

```

```

#cross-references MWID:8931567
#accession S04713
#molecule_type DNA
#residues 1-1460 #label CHE

CLASSIFICATION #superfamily herpesvirus immediate-early protein IE175
KEYWORDS DNA binding; early protein; transcription regulation
SUMMARY #length 1460 #molecular-weight 152851 #checksum 5209

Query Match 6.1%; Score 110; DB 1; Length 1460;
Best Local Similarity 40.0%; Pred. No. 5,996-02;
Matches 20; Conservative 7; Mismatches 21; Indels 2; Gaps 2;

Db 308 RRR-RARRPRREGRGTRPRRGGA-PLQRQPRRRRAGEALRRGRGFS 355
    III I III I : II : : : III : I :
QY 5 RRRYRRRPRSHLQILRRRLPWLPHRRYRWRRKNGIENRLSRFG 54

RESULT 14
ENTRY
TITLE
ORGANISM
DATE
REFERENCES
#authors

S00180 #type complete
spermatid protein sl - smaller spotted catshark
#formal_name Scyllorhinus canicula #common_name smaller
spotted catshark, smaller spotted dogfish
30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change
23-Feb-1997
S00180
S00180
Chauviere, M.; Martinge, A.; Briand, G.; Sautiere, P.;
Chevaillier, P.

```

REFERENCE  
#authors  
Sautiere, P.  
Chevallier, P.  
Martinge, A.; Briand, G.; Sautiere, P.;  
Chauviere, M.

```

#cross-references #MUID:8805030
#accession S00180
#molecule_type protein
#residues 1-87 ##label CHA
CLASSIFICATION superfamily sperm histone
KEYWORDS DNA binding; nucleus; phosphoprotein; sperm
FEATURE
3,55 #binding_site phosphate (Ser) (covalent) #status
experimental
SUMMARY
#length 87 #molecular-weight 11179 #checksum 9188
Query Match 6.0%; Score 108.; DB 2; Length 87;
Best Local Similarity 29.8%; Pred. No. 1,08e-01;
Matches 17; Conservative 14; Mismatches 22; Indels 4; Gaps 4;
Db 2 KSIYRNRGRPRRYGRMRKTKCR-RKGR-RISRPRHTYRRRY-RKRVHLKRRS 55
: ||| ||| : : : : : : : : : : : : : : : : : : : : : : : :
QY 5 RRRYRRRRRHPSHLGOILR-RPMVLVHPRHRYRWRRKNGIENTRLRSFGYTVKRT 60

```

```

RESULT 15
ENTRY
TITLE
ORGANISM
DATE
ACCESSIONS
REFERENCE
#authors
C71159 #type complete
hypothetical protein PH0472 - Pyrococcus horikoshii
#formal_name Pyrococcus horikoshii
14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change
14-Aug-1998
C71159
AT1000
Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.;
Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.;
Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.;
Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.;
Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi,
A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.;
Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.
DNA Res. (1998) 5:55-76
Complete sequence and gene organization of the genome of a
hyper-thermophilic archaeobacterium, Pyrococcus horikoshii
OT3.
#cross-references MUID:98344137

```

Result No.	Query		DB	ID	Description	Pred. No.	
	Score	Match Length					
1	153	8.5	77	1	PRT2_SEPOF	SPERMATID-SPECIFIC PRO	8.86e-10
2	146	8.1	78	1	PRT1_NOTTO	SPERMATID-SPECIFIC PRO	1.43e-08
3	143	7.9	60	1	HSP1_NOTTO	SPERM PROTAMINE P1.	4.63e-08
4	139	7.7	63	1	HSP1_DROAU	SPERM PROTAMINE P1.	2.18e-07
5	137	7.6	59	1	HSP1_PHACI	SPERM PROTAMINE P1.	4.71e-07
6	131	7.3	60	1	HSP1_DASVI	SPERM PROTAMINE P1.	4.56e-06
7	131	7.3	61	1	HSP1_SARHA	SPERM PROTAMINE P1.	4.56e-06
8	130	7.2	61	1	HSP1_MACEU	SPERM PROTAMINE P1.	6.63e-06
9	130	7.2	61	1	HSP1_PARB1	SPERM PROTAMINE P1.	6.63e-06
10	130	7.2	61	1	HSP1_TRIVU	SPERM PROTAMINE P1.	6.63e-06
11	130	7.2	62	1	HSP1_MURLO	SPERM PROTAMINE P1.	6.63e-06
12	130	7.2	62	1	HSP1_DASRO	SPERM PROTAMINE P1.	6.63e-06
13	129	7.1	60	1	HSP1_MACGI	SPERM PROTAMINE P1.	9.62e-06
14	129	7.1	61	1	HSP1_MACRG	SPERM PROTAMINE P1.	9.62e-06
15	129	7.1	61	1	HSP1_ANTLA	SPERM PROTAMINE P1.	9.62e-06
16	128	7.1	63	1	HSP1_ANTST	SPERM PROTAMINE P1.	1.39e-05
17	126	7.0	57	1	HSP1_DIDMA	SPERM PROTAMINE P1.	2.91e-05
18	127	7.0	60	1	HSP1_MACAG	SPERM PROTAMINE P1.	2.02e-05
19	126	7.0	68	1	HSP1_PESCU	SPERM PROTAMINE P1.	2.91e-05
20	125	6.9	59	1	HSP1_MACRU	SPERM PROTAMINE P1.	4.20e-05
21	125	6.9	60	1	HSP1_CAEFU	SPERM PROTAMINE P1.	4.20e-05
22	124	6.9	61	1	HSP1_ANTSW	SPERM PROTAMINE P1.	6.06e-05
23	113	6.3	56	1	HSP_COTJA	SPERM HISTONE (PROTAMI	3.01e-03

3A	MARIN-PONTHIEU A., WOUTERS-TYROU D., BELAICHE D., SAUTIERE P.,	
3B	SCHINDLER P., VAN DORSSELAER A.;	
PI	"Cuttlefish sperm protamines 1. Amino acid sequences of two distinct	
RT	variants.";	
RT	EUR. J. BIOCHEM. 195;611-619(1991).	
CC	- - FUNCTION: CUTTLEFISH SPERMIOGENESIS IS CHARACTERIZED BY A DOUBLE	
CC	NUCLEAR PROTEIN TRANSITION: HISTONES -> SPERMATID-SPECIFIC	
CC	PROTEINS (T1/T2) -> PROTAMINES (SP1/SP2). THE PROTAMINES COMPACT	
CC	SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.	
CC	- - SUBCELLULAR LOCATION: NUCLEAR.	
CC	- - TISSUE SPECIFICITY: TESTIS.	
CC	- - DEVELOPMENTAL STAGE: SPERMIOGENESIS.	
CC	- - PTM: PHOSPHORYLATION OCCURS AT DIFFERENT DEGREES. THE	
CC	TRIPHOSPHORYLATED FORM MAY BE PREDOMINANT IN T2. SP2 APPEARS TO BE	
CC	PHOSPHORYLATED IN ELONGATED SPERMATIDS, BUT DEPHOSPHORYLATED IN	
CC	MATURE SPERM CELLS.	
CC	- - SIMILARITY: HIGH, TO S.OFICINALIS SPERMATID-SPECIFIC PROTEIN T1.	
DR	PIR; S14086; S14086.	
DR	PIR; B40973; B40973.	
KW	CHROMOSOMAL PROTEIN; NUCLEOSOME CORE; SPERMATOGENESIS; DNA-BINDING;	
KW	TESTIS; DNA CONDENSATION; NUCLEAR PROTEIN; PHOSPHORYLATION;	
KW	MULTIGENE FAMILY.	
FT	CHAIN 1; 77 SPERMATID-SPECIFIC PROTEIN T2.	
FT	CHAIN 22; 77 SPERM PROTAMINE SP2.	
FT	DOMAIN 1; 21 HYDROPHOBIC.	



CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; L35449; G598111; -  
 DR PROSITE; PS00048; PROTAMINE.P1; 1.  
 DR PFAM; PF00260; PROTAMINE.P1; 1.  
 KW CHROMOSOMAL PROTEIN; NUCLEOSOME CORE; SPERMATOGENESIS; DNA-BINDING;  
 FT TESTIS; DNA CONDENSATION; NUCLEAR PROTEIN.  
 FT INIT\_MET 0 0 BY SIMILARITY.  
 SQ SEQUENCE 63 AA; 8860 MW; 5645E4EF CRC32;

Query Match 7.7%; Score 139; DB 1; Length 63;  
 Best Local Similarity 42.6%; Pred. No. 2.18e-07; Indels 4; Gaps 4;  
 Matches 23; Conservative 11; Mismatches 16;

Db 2 RYRHSRSRSRY-RR-RRRRLLNRRRRYRRRRRRRRRRRRRRRR-RGYSRRR 52  
 ||||| : : : : : ||| : : : : : ||| : : : : :  
 QY 7 RYRRR-RHRPRSHLQILRRPWLHPHRYWRKNGINFNLSRTFGYTVKR 59

RESULT 5  
 ID HSP1\_PHACI STANDARD; PRT; 59 AA.

AC P79990;  
 DT 13-JUL-1998 (REL. 36, CREATED)  
 DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)  
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
 DE SPERM PROTAMINE P1.  
 GN PRM1.  
 OS PHASCOLARCTOS CINEREUS (KOALA).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; METATHERIA;  
 OC DIPROTODONTIA; PHASCOLARCTIDAE; PHASCOLARCTOS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA RETIEF J.D., KRAJEWSKI C., WESTERMAN M., WINKFEIN R.J., DIXON G.H.;  
 RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF  
 CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPLEX  
 CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.  
 CC -!- TISSUE SPECIFICITY: TESTIS.  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; U87789; G1850518; -  
 DR PROSITE; PS00048; PROTAMINE.P1; 1.  
 DR PFAM; PF00260; PROTAMINE.P1; 1.  
 KW CHROMOSOMAL PROTEIN; NUCLEOSOME CORE; SPERMATOGENESIS; DNA-BINDING;  
 FT TESTIS; DNA CONDENSATION; NUCLEAR PROTEIN.  
 FT INIT\_MET 0 0 BY SIMILARITY.  
 SQ SEQUENCE 59 AA; 8271 MW; 98F41F75 CRC32;

Query Match 7.6%; Score 137; DB 1; Length 59;  
 Best Local Similarity 40.8%; Pred. No. 4.71e-07;  
 Matches 20; Conservative 10; Mismatches 18; Indels 1; Gaps 1;

Db 2 RYRHSRSRSRYRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRGY 49  
 ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 QY 7 RYRRRRRPRSHLQILRRPWLHPHRYWRKNGINFNLSRTFGY 55

RESULT 6  
 ID HSP1\_DASVI STANDARD; PRT; 60 AA.

AC P42135; P42133;  
 DT 01-NOV-1995 (REL. 32, CREATED)  
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE SPERM PROTAMINE P1.  
 GN PRM1.  
 OS DASYURUS VIVERRINUS (SOUTHEASTERN QUOLL), AND DASYURUS HALLUCATUS.  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; METATHERIA;  
 OC DASYUROMORPHA; DASYURIDAE; DASYURUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-SPERM;  
 RX MEDLINE; 95215351.  
 RA RETIEF J.D., KRAJEWSKI C., WESTERMAN M., WINKFEIN R.J., DIXON G.H.;  
 RL "Molecular phylogeny and evolution of marsupial protamine P1 genes.";  
 RL PROC. R. SOC. LOND., B, BIOL. SCI. 259:7-14(1995).

CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF  
 CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPLEX  
 CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -!- TISSUE SPECIFICITY: TESTIS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; L35340; G598103; -  
 DR EMBL; L35341; G598105; -  
 DR PROSITE; PS00048; PROTAMINE.P1; 1.  
 DR PFAM; PF00260; PROTAMINE.P1; 1.  
 KW CHROMOSOMAL PROTEIN; NUCLEOSOME CORE; SPERMATOGENESIS; DNA-BINDING;  
 KW TESTIS; DNA CONDENSATION; NUCLEAR PROTEIN.  
 FT INIT\_MET 0 0 BY SIMILARITY.  
 SQ SEQUENCE 60 AA; 8246 MW; D670FAB3 CRC32;

Query Match 7.3%; Score 131; DB 1; Length 60;  
 Best Local Similarity 38.9%; Pred. No. 4.56e-06;  
 Matches 21; Conservative 14; Mismatches 15; Indels 4; Gaps 4;

Lb 2 RYRHSRSRSRYRR 52  
 ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 QY 7 RYRRR-RHRPRSHLQILRRPWLHPHRYWRKNGINFNLSRTFGYTVKR 59

RESULT 7

ID HSP1\_SARHA STANDARD; PRT; 61 AA.  
 AC P42151;  
 DT 01-NOV-1995 (REL. 32, CREATED)  
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
 DE SPERM PROTAMINE P1.  
 GN PRM1.  
 OS SARCOPHILUS HARRISII (TASMANIAN DEVIL), AND DASYURUS MACULATUS.  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; METATHERIA;  
 OC DASYUROMORPHA; DASYURIDAE; SARCOPHILUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-S.HARRISII; TISSUE-SPERM;  
 RX MEDLINE; 95215351.  
 RA RETIEF J.D., KRAJEWSKI C., WESTERMAN M., WINKFEIN R.J., DIXON G.H.;  
 RL "Molecular phylogeny and evolution of marsupial protamine P1 genes.";  
 RL PROC. R. SOC. LOND., B, BIOL. SCI. 259:7-14(1995).

[2]

RP SEQUENCE FROM N.A.

RC SPECIES-D.MACULATUS;

RA KRAJEWSKI C., YOUNG J., BUCKLEY L., WOOLLEY P.A., WESTERMAN M.;

RL "Reconstructing the taxonomic radiation of dasyurine marsupials with

cytochrome b, 12S rRNA, and protamine P1 gene trees.";

J. MAMMAL. EVOL. 4:217-236(1997).



```

RL PROC. R. SOC. LOND., B. BIOL. SCI. 259:7-14(1995).
CC -!- FUNCTION: PROTAMINE SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: TESTIS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L32744; G548332;
DR PROSITE; PS00048; PROTAMINE_P1; 1.
DR PFAM; PF00260; Protamine_P1; 1.
DR CHROMOSOMAL PROTEIN; NUCLEOSOME CORE; SPERMATOGENESIS; DNA-BINDING;
KW TESTIS; DNA CONDENSATION; NUCLEAR PROTEIN.
FT INIT_MET 0 0 BY SIMILARITY.
SQ SEQUENCE 61 AA; 8571 MW; 4380E49A CRC32;

Query Match 7.2%; Score 130; DB 1; Length 61;
Best Local Similarity 45.8%; Pred. No. 6.63e-06;
Matches 22; Conservative 8; Mismatches 16; Indels 2; Gaps 2;

Db 10 SRSRYRRRRRRSSRY-RSRRRRYRRRRRR-RGRRRRGGYRRYSR 55
QY 4 PRRYRRRRRSHLQILRRRPLVPHRYRRWRKNGIENFRLSR 51

RESULT 11
ID HSP1_MURLO STANDARD; PRT; 62 AA.
AC P42140; P42150; P42154;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE SPERM PROTAMINE P1.
GN PM1.
OS MUREXIA LONGICAUDATA,
OS PHASCOGAE TAPOTATA,
OS SMINTHOPSIS CRASSICAUDATA,
OS MYRMECOBIUS FASCIATUS (NUMBAT), AND
OS THYLACINUS CYNOCEPHALUS (TASMANIAN WOLF).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; METATHERIA;
OC DASYUROMORPHIA; DASYURIDAE; MUREXIA.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SPERM;
RX MEDLINE; 95215351.
RA RETIEF J.D., KRAJEWSKI C., WESTERMAN M., WINKFEIN R.J., DIXON G.H.;
RT "Molecular phylogeny and evolution of marsupial protamine P1 genes.";
RL PROC. R. SOC. LOND., B. BIOL. SCI. 259:7-14(1995).
CC -!- FUNCTION: PROTAMINE SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: TESTIS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L32744; G548332;
DR PROSITE; PS00048; PROTAMINE_P1; 1.
DR PFAM; PF00260; Protamine_P1; 1.
DR CHROMOSOMAL PROTEIN; NUCLEOSOME CORE; SPERMATOGENESIS; DNA-BINDING;
KW TESTIS; DNA CONDENSATION; NUCLEAR PROTEIN.
FT INIT_MET 0 0 BY SIMILARITY.
SQ SEQUENCE 61 AA; 8571 MW; 4380E49A CRC32;

Query Match 7.2%; Score 130; DB 1; Length 61;
Best Local Similarity 45.8%; Pred. No. 6.63e-06;
Matches 22; Conservative 8; Mismatches 16; Indels 2; Gaps 2;

Db 10 SRSRYRRRRRRSSRY-RSRRRRYRRRRRR-RGRRRRGGYRRYSR 55
QY 4 PRRYRRRRRSHLQILRRRPLVPHRYRRWRKNGIENFRLSR 51

RESULT 11
ID HSP1_MURLO STANDARD; PRT; 62 AA.
AC P42140; P42150; P42154;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE SPERM PROTAMINE P1.
GN PM1.
OS MUREXIA LONGICAUDATA,
OS PHASCOGAE TAPOTATA,
OS SMINTHOPSIS CRASSICAUDATA,
OS MYRMECOBIUS FASCIATUS (NUMBAT), AND
OS THYLACINUS CYNOCEPHALUS (TASMANIAN WOLF).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; METATHERIA;
OC DASYUROMORPHIA; DASYURIDAE; MUREXIA.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SPERM;
RX MEDLINE; 95215351.
RA RETIEF J.D., KRAJEWSKI C., WESTERMAN M., WINKFEIN R.J., DIXON G.H.;
RT "Molecular phylogeny and evolution of marsupial protamine P1 genes.";
RL PROC. R. SOC. LOND., B. BIOL. SCI. 259:7-14(1995).
CC -!- FUNCTION: PROTAMINE SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: TESTIS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L32744; G548332;
DR PROSITE; PS00048; PROTAMINE_P1; 1.
DR PFAM; PF00260; Protamine_P1; 1.
DR CHROMOSOMAL PROTEIN; NUCLEOSOME CORE; SPERMATOGENESIS; DNA-BINDING;
KW TESTIS; DNA CONDENSATION; NUCLEAR PROTEIN.
FT INIT_MET 0 0 BY SIMILARITY.
SQ SEQUENCE 61 AA; 8571 MW; 4380E49A CRC32;

Query Match 7.2%; Score 130; DB 1; Length 61;
Best Local Similarity 45.8%; Pred. No. 6.63e-06;
Matches 22; Conservative 8; Mismatches 16; Indels 2; Gaps 2;

Db 10 SRSRYRRRRRRSSRY-RSRRRRYRRRRRR-RGRRRRGGYRRYSR 55
QY 4 PRRYRRRRRSHLQILRRRPLVPHRYRRWRKNGIENFRLSR 51

```

```

CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L32744; G548332;
DR PROSITE; PS00048; PROTAMINE_P1; 1.
DR PFAM; PF00260; Protamine_P1; 1.
DR CHROMOSOMAL PROTEIN; NUCLEOSOME CORE; SPERMATOGENESIS; DNA-BINDING;
KW TESTIS; DNA CONDENSATION; NUCLEAR PROTEIN.
FT INIT_MET 0 0 BY SIMILARITY.
SQ SEQUENCE 62 AA; 8566 MW; BEF8BCB4 CRC32;

Query Match 7.2%; Score 130; DB 1; Length 62;
Best Local Similarity 37.0%; Pred. No. 6.63e-06;
Matches 20; Conservative 15; Mismatches 16; Indels 3; Gaps 3;

Db 7 RYRRHRSRSRYRRRRRRSHHNRRTYRRSRHSR-RRR-GRRRGYRRR 53
QY 7 RYRRR-RHRPRSHLQILRRRPLVPHRYRRWRKNGIENFRLSR 59

RESULT 12
ID HSP1_DASRO STANDARD; PRT; 62 AA.
AC P42134; P42144; P42149;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE SPERM PROTAMINE P1.
GN PM1.
OS DASYKALUTA ROSAMONDAE, PARANTECHINUS APICALIS, AND
OS PSEUDANTECHINUS MACDONNELLENSIS.
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; METATHERIA;
OC DASYUROMORPHIA; DASYURIDAE; DASYKALUTA.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SPERM;
RX MEDLINE; 95215351.
RA RETIEF J.D., KRAJEWSKI C., WESTERMAN M., WINKFEIN R.J., DIXON G.H.;
RT "Molecular phylogeny and evolution of marsupial protamine P1 genes.";
RL PROC. R. SOC. LOND., B. BIOL. SCI. 259:7-14(1995).
CC -!- FUNCTION: PROTAMINE SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: TESTIS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L32744; G548332;
DR PROSITE; PS00048; PROTAMINE_P1; 1.
DR PFAM; PF00260; Protamine_P1; 1.
DR CHROMOSOMAL PROTEIN; NUCLEOSOME CORE; SPERMATOGENESIS; DNA-BINDING;
KW TESTIS; DNA CONDENSATION; NUCLEAR PROTEIN.
FT INIT_MET 0 0 BY SIMILARITY.
SQ SEQUENCE 62 AA; 8565 MW; DAF6508 CRC32;

Query Match 7.2%; Score 130; DB 1; Length 62;
Best Local Similarity 37.0%; Pred. No. 6.63e-06;
Matches 20; Conservative 15; Mismatches 16; Indels 3; Gaps 3;

Db 7 RYRRHRSRSRYRRRRRRSHHNRRTYRRSRHSR-RRR-GRRRGYRRR 53
QY 7 RYRRR-RHRPRSHLQILRRRPLVPHRYRRWRKNGIENFRLSR 59

```



[illegible]

CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- TISSUE SPECIFICITY: TESTIS.

---

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).

---

CC EMBL; L353329; G598421;  
DR EMBL; L353328; G598417;  
DR PROSITE; PS00048; PROTAMINE\_P1; 1.  
DR PFAM; PF00260; protamine\_P1; 1.  
KW CHROMOSOMAL PROTEIN; NUCLEOSOME CORE; SPERMATOGENESIS; DNA-BINDING;  
KW TESTIS; DNA CONDENSATION; NUCLEAR PROTEIN.  
FT INIT\_MET 0 0 BY SIMILARITY  
SQ SEQUENCE 61 AA; 8571 MW; 73734DB3 CRC32;

Query Match 7.1%; Score 129; DB 1; Length 61;  
Best Local Similarity 38.9%; Pred. No. 9.62e-06;  
Matches 21; Conservative 12; Mismatches 17; Indels 4; Gaps 4;

Db 2 RYHRSRSGSRVRRRRRRSRYRRRRRGRRRRR-RGR-RRR-GYSRRR 52  
|||:|:||:|:||||:||||:|:||||:|:|:|:|:  
QY 7 RYRRRRHRPSHLGQLLRRLPVLPHRHRYR-WRKNGINFTRLSRTFGYTVR 59

---

RESULT 15  
ID HSP1\_ANTLA STANDARD; PRT; 61 AA.  
AC O18745;  
DT 15-JUL-1998 (REL. 36, CREATED)  
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE SPERM PROTAMINE P1.  
GN PRM1.  
OS ANTECHINOMYS LANIGER.  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; METATHERIA;  
OC DASYUROMORPHA; DASYURIDAE; ANTECHINOMYS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 97446280.  
RA KRAJEWSKI C., BLACKET M., BUCKLEY L., WESTERMAN M.;  
RT "A multigene assessment of phylogenetic relationships within the  
RT dasyurid marsupial subfamily Sminthopsinae";  
RL MOL. PHYLOGENET. EVOL. 8:236-248(1997).  
CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF  
CC SPERM DURING THE HAPOLOID PHASE OF SPERMATOGENESIS. THEY COMPACT  
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.

---

CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- TISSUE SPECIFICITY: TESTIS.

---

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).

---

CC EMBL; AF001587; G2330919;  
DR PROSITE; PS00048; PROTAMINE\_P1; 1.  
DR PFAM; PF00260; protamine\_P1; 1.  
KW CHROMOSOMAL PROTEIN; NUCLEOSOME CORE; SPERMATOGENESIS; DNA-BINDING;  
KW TESTIS; DNA CONDENSATION; NUCLEAR PROTEIN.  
FT INIT\_MET 0 0 BY SIMILARITY  
SQ SEQUENCE 61 AA; 8409 MW; 106BF7EF CRC32;

Query Match 7.1%; Score 129; DB 1; Length 61;  
Best Local Similarity 38.9%; Pred. No. 9.62e-06;  
Matches 21; Conservative 14; Mismatches 15; Indels 4; Gaps 4;

Db 2 RYRRSRSRSRSRSRSRSRH-HNRRTYRRSRHSP-RRR-GRRGYSRRR 52  
QY 7 RYRRSRSRSRSRSRSRH-HNRRTYRRSRHSP-RRR-GRRGYSRRR 59

Search completed: Wed Dec 22 10:14:37 1999  
Job time : 10 secs.

\*\*\*\*\*  
W O R L D  
\*\*\*\*\*  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Dec 22 10:19:10 1999; MasPar time 12.17 Seconds  
Tabular output not generated. 103.094 Million cell updates/sec

Title: >US-09-209-961-7  
Perfect Score: (1-59) from US09209961.pep  
Sequence: 1 MKCTLVQSRFCIFPLTFKS.....TRIPLSNKVLTAVDRLRCP 59

Scoring table: PAM 150  
Gap 11

Searched: 170751 seqs, 21266508 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq35  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38

Statistics: Mean 25.947; Variance 100.993; scale 0.257

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	79	17.8	112	33	W60650 Human C-C chemokine D	2.34e+01
2	77	17.3	91	20	R97665 RANTES	3.41e+01
3	77	17.3	91	13	R70802 RANTES protein.	3.41e+01
4	77	17.3	154	31	W37864 Human protein compri	3.41e+01
5	77	17.3	263	26	R63905 Type I ribosome-inact	3.41e+01
6	77	17.3	263	1	R74179 Type I ribosome-inact	3.41e+01
7	77	17.3	263	7	R37293 Plant type I RFP Womo	3.41e+01
8	76	17.1	310	22	W20383 H. pylori cytoplasmic	4.11e+01
9	76	17.1	326	39	W83311 Human AL adenosine re	4.11e+01
10	76	17.1	326	16	R87655 Human adenosine recep	4.11e+01
11	76	17.1	326	18	R93989 Human ventricle AL ad	4.11e+01
12	76	17.1	326	9	R47386 Human AL adenosine re	4.11e+01
13	76	17.1	326	8	R41524 Human AL adenosine re	4.11e+01
14	76	17.1	326	9	R45744 Human AL adenosine re	4.11e+01
15	76	17.1	326	15	R84192 Human AL adenosine re	4.11e+01
16	76	17.1	446	22	W20836 H. pylori cytoplasmic	4.11e+01

17	76	17.1	648	23	W24473	Herbicide-sensitive a	4.11e+01
18	76	17.1	679	18	R96806	Human prostate transg	4.11e+01
19	76	17.1	679	8	R39388	Human prostate transg	4.11e+01
20	75	16.9	318	20	W06405	Verotoxinigenic E. coli	4.95e+01
21	75	16.9	326	20	W06409	Histidine tagged vero	4.95e+01
22	75	16.9	473	19	R7244	Virulence gene cluste	4.95e+01
23	74	16.6	227	22	W20816	H. pylori cytoplasmic	5.96e+01
24	72	16.2	64	10	R50301	Drosophila anti-bacte	8.62e+01
25	72	16.2	64	10	R50293	Anti-bacterial glycop	8.62e+01
26	72	16.2	275	38	W83380	Human uncoupling prot	8.62e+01
27	72	16.2	312	38	W83379	Human uncoupling prot	8.62e+01
28	72	16.2	312	35	W68197	Human uncoupling prot	8.62e+01
29	72	16.2	345	10	R54215	L.lactis branched ami	8.62e+01
30	71	16.0	90	39	W82720	Human S1SD protein.	1.03e+02
31	71	16.0	149	38	W82648	Ehrlichia sp. extende	1.03e+02
32	71	16.0	275	37	W81591	Human uncoupling prot	1.03e+02
33	71	16.0	331	39	W89491	Hepatitis G virus ant	1.03e+02
34	71	16.0	337	39	W89489	Hepatitis G virus ant	1.03e+02
35	71	16.0	350	12	R67001	Cannabinoid receptor	1.03e+02
36	71	16.0	2723	31	W56448	Fragment HGJ1775 of a	1.03e+02
37	71	16.0	2873	39	W89452	Hepatitis G virus PNF	1.03e+02
38	71	16.0	2873	36	W80148	HGV-PNF2161 variant s	1.03e+02
39	71	16.0	2910	39	W89458	Hepatitis G virus var	1.03e+02
40	71	16.0	2910	17	R87566	Hepatitis G virus clo	1.03e+02
41	71	16.0	2969	31	W56446	Fragment HGJ2141 of a	1.03e+02
42	71	16.0	2969	31	W56442	Fragment HGJ1737 of a	1.03e+02
43	71	16.0	2969	31	W56449	Fragment K3732(2V) of	1.03e+02
44	71	16.0	2969	31	W56443	Fragment HGJ1741 of a	1.03e+02
45	71	16.0	3011	16	R95020	Hepatitis GB virus (H	1.03e+02

ALIGNMENTS

RESULT 1

ID W60650 standard; Protein: 112 AA.

AC W60650;

DT 12-CCT-1998 (first entry)

DE Human C-C chemokine DGWCC.

KW DGWCC; DNAX groin wound expressed CC chemokine; cytokine; human;

WV immune system; cancer; cell proliferation; therapy; diagnosis.

OS Homo sapiens.

FH Key 'Location/Qualifiers

FT Peptide 1..24

FT Protein /label= Sig\_peptide

FT /label= Mat\_protein

FT /note= 'Claim 2'

PN W09823750-A2.

PD 04-JUN-1998.

PR 26-NOV-1997; U21092.

PR 05-DEC-1996; US-761071.

PR 27-NOV-1996; US-031805.

PA (SCHE ) SCHERING CORP.

PI Hedrick JA, Morales J, Vicari A, Zlotnik A;

DR WPI: 98-322730/28.

DR N-PSDB: V38294.

FT DVic-1 and DGWCC chemokines - useful for developing products for

FT treating abnormal physiology or development, e.g. cancerous or

FT degenerative conditions

FT Claim 2; Page 62; 71pp; English.

CC This polypeptide comprises novel human DNAX groin wound expressed

CC chemokine (DGWCC), the mature portion of which is claimed. The

CC DGWCC amino acid sequence was deduced from a cDNA clone (see V38294).

CC also claimed is novel human DNAX Vic-1 (DVic-1) (see W09823750-A2).

CC polypeptide, as well as expression vectors and host cells. DGWCC

CC or DVic-1 play a role in the regulation or development of neuronal

CC or haematopoietic cells, e.g. lymphoid cells, which affect

CC immunological responses. They can be used in the treatment of

CC conditions associated with abnormal physiology or development,

CC including abnormal proliferation, e.g. cancerous conditions or

CC degenerative conditions. Abnormal proliferation, regeneration,

CC degeneration, and atrophy may be modulated by appropriate

CC therapeutic treatment using products of the invention. The products

CC can also be used for detection, diagnosis and drug screening.  
SQ Sequence 112 AA;

Query Match 17.8%; Score 79; DB 33; Length 112;  
Best Local Similarity 37.8%; Pred. No. 2.34e+01;  
Matches 17; Conservative 10; Mismatches 14; Indels 4; Gaps 4;

Db 7 fcslllllllspdaafllpbtacc-tqlyrkplskllrv 50  
QY 11 FC-IFPLTFKSSAPRK-FL-TNWTGCCFATVTRIPLSNKLTVAV 52

## RESULT 2

ID R97665 standard; Protein; 91 AA.

AC R97665; 14-FEB-1997 (first entry)  
DE RANTES.  
KW Human; RANTES; Raised on Activation, normal T cell derived and secreted;  
KW bone marrow; antagonist; MIP-1alpha; asthma; allergic rhinitis;  
KW atopic dermatitis; atheroma; atherosclerosis; rheumatoid arthritis;  
KW inflammation; chemotaxis.  
OS Homo sapiens.

FS Key Location/Qualifiers

FT peptide 1..23 /note= "Signal peptide"

FT protein 24..91 /note= "RANTES, claim 3"

FT WO9617935-A2.

PN 13-JUN-1996.

PD 07-DEC-1995; G02861.

PR 08-DEC-1994; GB-024835.

PR 16-JUN-1995; GB-012319.

PA (GLAX) GLAXO GROUP LTD.

PI Proudfoot AEI, Wells TNC;

DR WPI; 96-287180/29.

DR N-PSDB; T30329.

PT Polypeptide functioning as antagonist to RANTES or to MIP-1-alpha  
PT used in the treatment of e.g. asthma, allergic rhinitis,  
PT atheroma/atherosclerosis or rheumatoid arthritis  
PS Example; Fig 1; 33pp; English.  
CC This sequence represents full length human RANTES (Raised on Activation,  
CC normal T cell derived and secreted). The DNA encoding this sequence was  
CC cloned from a human bone marrow lambda gt10 cDNA library using the primer  
CC sequences given in T30330-33. RANTES was used in the design of  
CC peptides which act as antagonists to RANTES or to MIP-1alpha (see  
CC also R97666-68). These peptides may be used in the treatment of  
CC asthma, allergic rhinitis, atopic dermatitis, atheroma/atherosclerosis  
CC or rheumatoid arthritis by inhibiting or reducing inflammation mediated  
CC by RANTES or MIP-1alpha. They are also useful in studying RANTES  
CC induced chemotaxis, mobilisation of Ca2+ and receptor binding.  
SQ Sequence 91 AA;

Query Match 17.3%; Score 77; DB 20; Length 91;  
Best Local Similarity 38.5%; Pred. No. 3.41e+01;  
Matches 10; Conservative 9; Mismatches 6; Indels 1; Gaps 1;

Db 19 apaspsysdttppccfayiar-plp 43

QY 20 SSASPRKFLNVTGCCFATVTRIPLS 45

## RESULT 3

ID R70802 standard; Protein; 91 AA.

AC R70802;

DE 29-AUG-1995 (first entry)

DE RANTES Protein.

KW RANTES; T-cell; T-lymphocyte; heparanase; heparin; heparan sulfate;  
KW arthritis; restenosis; cancer; wound healing.

OS Homo sapiens.

PN WO9504158-A.

PD 09-FEB-1995.

PF 26-JUL-1994; U08207.

PR 24-JUL-1993; US-099866.

PR 13-OCT-1993; US-136117.

PA (UPJO) UPJOHN CO.

PI Hoogwerf AJ, Ledbetter SR;

DR WPI; 95-082239/11.

DR N-PSDB; Q85372.

PT Screening for cpds. with anti-heparanase activity - by detecting

PT inhibition of heparin or heparan sulphate degradation, cancer.

PT potentially useful for treating arthritis, restenosis, cancer.

PS Claim 13; Page 51; 60pp; English.

CC Activated with transglutaminase, are given in R70786-802. Most

CC purified heparanases, prepared under reducing conditions and

CC are prepared by reverse transcription of mRNA from activated human

CC leukocytes, then cloning of the cDNA into pVL1392 baculovirus

CC vector, and expression in Sf9 cells in the presence of reduced

CC glutathione and dithiothreitol.

SQ Sequence 91 AA;

Query Match 17.3%; Score 77; DB 13; Length 91;

Best Local Similarity 38.5%; Pred. No. 3.41e+01;

Matches 10; Conservative 9; Mismatches 6; Indels 1; Gaps 1;

Db 19 apaspsysdttppccfayiar-plp 43

QY 20 SSASPRKFLNVTGCCFATVTRIPLS 45

## RESULT 4

ID W37864 standard; Protein; 154 AA.

AC W37864;

DT 10-AUG-1998 (first entry)

DE Human protein comprising secretory signal amino acid sequence 1.

KW Human protein; secretory signal; nutritional source; cytokine;

KW immunity; haematopoiesis; activin; inhibin; tumour; chemotactic;

KW chemokine; thrombolytic; anti-inflammatory; inhibition;

KW fibrosarcoma cell.

OS Homo sapiens.

PN WO9811217-A2.

PD 19-MAR-1998.

PF 12-SEP-1997; J03239.

PR 13-SEP-1996; JP-243060.

PA (PROT-) PROTEGENE INC.

PA (SAGA) SAGAMI CHEM RES CENTRE.

PI Kato S, Kobayashi M, Sekine S, Yamaguchi T;

DR N-PSDB; V29031, V29032.

DR WPI; 98-207380/18.

PT Human proteins with secretory signal sequences - used to treat

PT immune deficiencies, infections, tumours, and haematopoietic

PT disorders, etc.

PS Claim 1; Pages 64-65; 131pp; English.

CC This is the amino acid sequence of a novel human protein comprising

CC a secretory signal isolated from fibrosarcoma cells. Its proteins

CC can be used as nutritional sources or supplements. The proteins may

CC also have cytokine functions, immune modulating functions,

CC haematopoiesis regulating activity, activin/inhibin regulating

CC activity, chemotactic/chemokinetic activity, haemostatic and

CC thrombolytic activity, receptor/ligand activity, anti-inflammatory

CC activity, tumour inhibition activity.

SQ Sequence 154 AA;

Query Match 17.3%; Score 77; DB 31; Length 154;

Best Local Similarity 38.5%; Pred. No. 3.41e+01;

Matches 10; Conservative 9; Mismatches 6; Indels 1; Gaps 1;

Db 19 apaspsysdttppccfayiar-plp 43

QY 20 SSASPRKFLNVTGCCFATVTRIPLS 45

## RESULT 5

ID R63905 standard; Protein; 263 AA.

AC R63905;

DT 28-JUL-1995 (first entry)

DE Type I ribosome-inactivating protein mormordin II.



RESULT	8
ID	W20383 standard; protein; 310 AA.
AC	W20383;
DE	14-JUL-1997 (first entry)
DE	H. pylori cytoplasmic protein, 29844512.aa.
DE	Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW	binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW	duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
KW	Helicobacter pylori.
OS	Helicobacter pylori.
FH	Key Location/Qualifiers
FT	misc_difference 299
FT	/note= "encoded by RCT"
PN	WO9640893-A1.
PN	19-DEC-1996.
PF	06-JUN-1996; U09122.
PF	07-JUN-1995; US-487032.
PR	01-APR-1996; US-630405.
PR	(ASTR ) ASTRA AB.
PA	Berglindh OT, Smith D, Mellgaard BL;
PI	WPI: 97-052306/05.
DR	N-PSDB; T67570.
DR	Helicobacter pylori nucleic acid sequences and related
PT	polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT	infection, and to detect Helicobacter
PS	Claim 61; Page 569; 1481pp; English.
PS	The present sequence is a H. pylori cytoplasmic protein.
CC	The protein may be used in a vaccine to prevent or treat H. pylori
CC	infection or to identify H. pylori polypeptide binding compounds,
CC	useful as potential H. pylori life cycle activators or inhibitors.
CC	The genomic sequence of H. pylori (ATCC 55679) was determined from
CC	overlapping contigs generated by mechanically shearing the bacterial
CC	DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC	and the predicted coding regions defined by computer evaluation. To
CC	identify likely H. pylori antigens for vaccine development, the amino
CC	acid sequences predicted from various ORF were analysed for significant
CC	homology to other known or exported membrane proteins. Having identified
CC	and determined the sequences of interest, particular regions can be
CC	isolated from H. pylori by PCR amplification for recombinant polypeptide
CC	production, e.g. in E. coli hosts.
SC	Sequence 310 AA;

```

Query Match      17.1%; Score 76; DB 22; Length 310;
Best Local Similarity 22.7%; Pred.No. 4.ille-01;
Matches 10; Conservative 15; Mismatches 1; Gaps 1;

Db 26 lkrslslsanhcvlsagygassaikxf-qeillgcviqspktnkl 68
QV 1 MKCTLVQSRFCFFITFTKSSASPREFLTNVTGCCFATVTRPL 44

```

RESULT	9	
OS	W82311 standard; Protein; 326 AA.	
ID	W82311;	
AC	15-MAR-1999 (first entry)	
DT	Human A1 adenosine receptor.	
DE	Adenosine receptor A1; human; cardioprotective; screening; assay;	
KW	agonist; antagonist; ischaemia; angina; myocardial infarction;	
KW	therapy.	
KW		
OS	Homo sapiens.	
PN	WO9850047-A1.	
PD	12-NOV-1998.	
PF	08-MAY-1998; U09031.	
PR	10-OCT-1997; US-061716.	
PR	09-MAY-1997; US-046030.	
PA	(UIPE-) UNIV PENNSYLVANIA.	
PI	Jacobson KA, Liang BT;	
DR	WPI: 99-034683/03.	
DR	N-PSDB: V73320.	
PT	Preventing or reducing ischaemic injury to heart muscle - by	
PT	activating adenosine A1 and A3 receptors and optionally also	
PT	antagonising A2a receptors, used for drug screening	
PS	Example 5; Fig 11A-B; 85pp; English.	

```

CC This is the amino acid sequence of human adenosine receptor A1.
CC In a claimed method, ischaemic damage to the heart is prevented or
CC reduced by administering: (i) an agonist that activates both
CC cardiac A1 and A3 adenosine receptors, optionally together with an
CC antagonist that inhibits activity of cardiac A2a adenosine
CC receptors; (ii) a binary conjugate that acts as A3 agonist and A2a
CC antagonist; and (iii) at least one each of A1-specific and
CC A3-specific agonists. The method is used to treat a patient having
CC chronic stable angina, unstable angina or post myocardial infarction
CC angina (claimed). Also claimed are recombinant cardiac myocytes
CC comprising a nucleic acid (see V73320-22) encoding the human A1,
CC A2a or A3 receptor (see W82311-13). The recombinant myocytes are
CC used in a claimed method for determining whether a test compound
CC exerts a cardioprotective effect.
SQ Sequence 326 AA;

Query Match 17.1% Score 76; DB 39; Length 326;
Best Local Similarity 28.9%; Pred. No. 4.11e+01;
Matches 13;...Conservative 13; Mismatches 16; Indels 3; Gaps 3;

Db| 111 |iplykmvtprraaavaiaagcwllsfvgltpmfgwnnlssavera 155
   |||||::||::||::||::||::||::||::||::||::||::||:
QY 14 FPLTFKSSASPRFLNTVTC-CFA-TVTRIPLSN-KVLTAVDRS 55

RESULT 10
ID R87655 standard; Protein; 326 AA.
AC R87655;
DE DT 19-JUN-1996 (first entry)
DE Human adenosine receptor A1 subtype.
KW Primer: probe; reverse transcription polymerase chain reaction; RT-PCR;
KW in situ hybridisation; adenosine receptor subtype; A3; A1; A2a; A2b;
KW human; eosinophil; antagonist; inhibition; adenylate cyclase;
KW cytochrome induced hypersensitivity; allergy; inflammatory disease;
KW asthma; rhinitis; serum sickness; atopic dermatitis; psoriasis; eczema;
KW ulcerative colitis; canine; ventricle; kidney.
```

RESULT .10  
ID R87655 standard: protein: 326 AA

AC	R97655;
DT	19-JUN-1996 (first entry)
DE	Human adenosine receptor A1 subtype.
KW	Primer; probe; reverse transcription polymerase chain reaction; RT-PCR;
KW	in situ hybridisation; adenosine receptor subtype; A3; A1; A2a; A2b;
KW	human; eosinophil; antagonist; inhibition; adenylate cyclase;
KW	cytokine induced hypersensitivity; allergy; inflammatory disease;
KW	asthma; rhinitis; serum sickness; atopic dermatitis; psoriasis; eczema;
KW	ulcerative colitis; canine; ventricle; kidney.
OS	Homo sapiens.
PN	G8288733-A.
ED	01-NOV-1995.
PF	19-APR-1995;. 007984.
PR	25-APR-1994; US-233009.
PA	(MERI ) MERCK & CO INC.
PI	Jacobson MA, Johnson RG,
DR	WPI.; 95-360715/47.
DR	N-PSDB: T07648.

Decreasing intracellular cAMP levels in eosinophil(s) - by  
contacting cell with antagonist of A3 subtype adenosine receptor.  
Disclosure; Fig 1; 12pp; English.  
This sequence represents the human adenosine receptor A1 subtype. RNA  
was extracted from homogenised human ventricle tissue and probed using  
synthetic probes based on the canine A1 sequence. The isolated human  
A1 clone is 92% identical to the canine A1 sequence. The isolated human  
sequence was amplified and probe fragments were used to screen a human  
cDNA library. The isolated sequence was used in the  
construction of the full length A1 cDNA. The full length cDNA may be  
used in the determination of expression of the A3 adenosine receptor by  
human eosinophils. This is used in the identification of antagonists of  
A3 subtype adenosine receptors. These antagonists prevent activation of  
eosinophils, esp. they prevent A3 inhibition of adenylate cyclase and  
cytokine induced hypersensitivity. They may be used for treating or  
preventing allergic or inflammatory diseases, e.g. asthma, rhinitis,  
serum sickness, atopic dermatitis, psoriasis, eczema, ulcerative colitis  
etc.

CC etc. 326 AA: 1  
SO Sequence

```

Query Match      ... 17.18; Score 76; DB 15; Length 326;
Best Local Similarity 28.98; pred. No. 4.11e+01;
Matches 13; Conservative 13; Mismatches 16; Indels 3; Gaps 3;

Db 111 lplrykmvncpraaavaiaagcwlsvfvgltpmfgwnnisavera 155
      :|||::|||::|||::|||::|||::|||::|||::|||::|||::
Qy 14 FPLTFKSSASPRFLNTVTC-CFA-TVTRIPLSN-KVLTAVDRS 55

```

```

RESULT 11
ID R39989 standard; Protein; 326 AA.
AC R39989;
DE 28-OCT-1996 (first entry)
DT Human ventricle A1 adenosine receptor.
KW Human ventricle A1 adenosine receptor; COS7; Xenopus oocyte;
KW ligand binding; ss.
OS Homo sapiens.
PN W09511681-A1.
PD 04-MAY-1995.
PF 26-OCT-1994; U12272.
PR 29-OCT-1993; US-145437.
PA (MERI) MERCK & CO INC.
PA (UYVI-) UNIV VIRGINIA PATENTS FOUND.
PI Doyle MP, Duling BR, Jacobson MA, Johnson RG, Linden JM;
DR WPI: 95-178635/23.
DR N-PSDB: T29929.
PT Use of xanthine derivs. as adenosine A3 receptor inhibitors - used
PT to block vasoconstriction, e.g. to treat myocardial ischaemia or
PT reperfusion injury, inflammation, auto-immune disorders, allergies,
PT brain arteriole constriction, etc.
PS Disclosure: Fig 1: 103pp; English.
CC Human ventricle A1 adenosine receptor has high affinity for
CC xanthine-agonists with saturated rings in the N6 position of the
CC adenine ring and xanthine-antagonists with saturated rings in the
CC C8 position. Recombinant receptor may be produced by subcloning
CC the cDNA into plasmid pSVL, plasmid-pCMV5, or plasmid prep.
CC Membranes prepared from transfected COS7 (ATCC CRL 1651) cells
CC were utilized for the determination of binding affinity,
CC selectivity and specificity of the human adenosine receptor for
CC various ligands. A transient expression system in Xenopus oocytes,
CC established by microinjection of in vitro transcribed mRNA from the
CC cloned cDNA, also allows measurement of the biological effects upon
CC activation of the expressed adenosine receptor with ligand binding.
SQ Sequence. 326 AA;

Query Match 11 17.1%; Score 76; DB 18; Length 326;
Best Local Similarity 28.9%; Pred. No. 4.11e+01;
Matches 13; Conservative 13; Mismatches 16; Indels 3; Gaps 3;

Db 111 lplrykmvtprraavagcwlsvfvgltmfgwnnlisavera 155
14 FPLTFKSSASPRKFLTNVTGC-CFA-TVTRIPLSN-KVLTAVDRS 55

RESULT 12
ID R47386 standard; Protein; 326 AA.
AC R47386;
DE 29-JUN-1994 (first entry)
DT Human A1 adenosine receptor.
KW G protein-coupled receptor; agonist; antagonist; sedative;
KW epilepsy; bronchoconstriction; hypotensive.
OS Homo sapiens.
PN W09325677-A.
PD 23-DEC-1993.
PF 11-JUN-1993; AU0277.
PR 12-JUN-1992; AU-002936.
PA (GARV-) GARVAN INST MEDICAL RES.
PI Furlong T, Pierce KD, Seibie L, Shine J, Townsend-Nicholson CA;
PI Seibie L;
DR WPI: 94-007535/01.
DR N-PSDB: Q54840.
PT DNA sequences encoding the human A1, A2a and A2b adenosine
PT receptors - for screening for anti-hypertensives, hypnotics,
PT anti-psychotics, bronchodilators and for recombinant receptor
PT prodn.
PS Disclosure: Fig 1: 26pp; English.
CC The sequence of the human A1 adenosine receptor was isolated from
CC a human hippocampal cDNA library by screening with specific
CC consensus oligonucleotides corresponding to the second extracellular
CC loop (Q54639) and to the third intracellular loop (Q54638). The

```

CC coding sequence is useful for recombinant production of the A1  
 CC receptor to facilitate agonist/antagonist screening. The amino acid  
 CC s-sequence R47386 was deduced from the cDNA sequence.

Query Match 11 17.1%; Score 76; DB 9; Length 326;  
 Best Local Similarity 28.9%; Pred. No. 4.11e+01;  
 Matches 13; Conservative 13; Mismatches 16; Indels 3; Gaps 3;

Db 111 lplrykmvtprraavagcwlsvfvgltmfgwnnlisavera 155  
 14 FPLTFKSSASPRKFLTNVTGC-CFA-TVTRIPLSN-KVLTAVDRS 55

#### RESULT 13

ID R41524 standard; Protein; 326 AA.

AC R41524;

DT 09-MAR-1994 (first entry)

DE Human A1 adenosine receptor.

KW Adenosine receptor; expression system; cardiovascular disorders;

KW renal disorders; neurological disorders; assay.

OS Homo sapiens.

PN GB2264948-A.

PD 15-SEP-1993.

PF 05-MAR-1993; 004582.

PR 13-MAR-1992; US-850701.

PR 13-MAR-1992; US-850702.

PR 13-MAR-1992; US-850707.

PR 15-JAN-1993; US-005945.

PA (MERI) MERCK & CO INC.

DR WPI: 93-290442/37.

DR N-PSDB: Q48414.

PT Human adenosine receptor proteins - and expression systems,

PT including cells for use in adenosine receptor binding assay

PS Claim 5; Figure 1; 65pp; English.

CC Purified adenosine receptor proteins can be used in assays to screen

CC for therapeutic human adenosine receptor agonists, antagonists or

CC binding enhancers. These agonists, antagonists and binding

CC enhancers may be used in the treatment of cardiovascular, renal and

CC neurological disorders.

SQ Sequence. 326 AA;

#### Query Match

11 17.1%; Score 76; DB 8; Length 326;

Best Local Similarity 28.9%; Pred. No. 4.11e+01;

Matches 13; Conservative 13; Mismatches 16; Indels 3; Gaps 3;

Db 111 lplrykmvtprraavagcwlsvfvgltmfgwnnlisavera 155

14 FPLTFKSSASPRKFLTNVTGC-CFA-TVTRIPLSN-KVLTAVDRS 55

#### RESULT 14

ID R45744 standard; Protein; 326 AA.

AC R45744;

DT 25-JUL-1994 (first entry)

DE Human A1 adenosine receptor.

KW Adenosine; adenylylate cyclase; screening; gene therapy; ischemia;

KW cystic fibrosis; RFLP.

OS Homo sapiens.

PN W09402605-A.

PD 03-FEB-1994.

PF 20-JUL-1993; U06782.

PR 22-JUL-1992; US-918314.

PA (UYDU-) UNIV DUKE.

PI Olah M, Ren H, Stiles GL;

DR WPI: 94-048863/06.

DR N-PSDB: Q55711.

PT DNA encoding the human A1 adenosine receptor - used for

PT developing agents for diagnosis or treatment of conditions

PT involving the receptor such as cystic fibrosis

PS Disclosure: Fig 2; 48pp; English.

CC The human A1 adenosine receptor genomic clone was isolated from a

CC human leukocyte genomic library in EMBL3 using bovine adenosine

CC receptor as probe. The genomic DNA was used as a probe to isolate  
 CC human A1 adenosine receptor cDNA clones from a human brain cDNA  
 CC library. The DNA encodes a protein which selectively binds  
 CC adenosine and inhibits adenylate cyclase. The protein may  
 CC be used to develop therapeutic agents to manipulate cardiac function,  
 CC protect against ischemia, regulate smooth muscle tone in blood  
 CC vessels or for treating cystic fibrosis or seizure activity.  
 CC See also R45742-3.  
 SQ Sequence 326 AA;

Query Match 17.1%; Score 76; DB 9; Length 326;  
 Best Local Similarity 28.9%; Pred. No. 4.11e+01;  
 Matches 13; Conservative 13; Mismatches 16; Indels 3; Gaps 3;

Db 111 iplyrkmvtprraavaagcwlsvfvgltptmfgwnnlsavera 155

QY 14 FPLTFKSSASPRFLTNVTC-CFA-TVTRIPLSN-KVLTAVDRS 55

# RESULT 15

ID R84192 standard; Protein; 326 AA.

AC R84192;

DT 25-MAR-1996 (first entry)

DE Human A1 adenosine receptor.

KW Adenosine receptor A1; adenosine receptor agonist;

KW autoimmune disease; therapy.

OS Homo sapiens.

PN GB228218-A.

PD 15-NOV-1995.

PF 01-MAY-1995; 008844.

PR 06-MAY-1994; US-239473.

PA (MERI ) MERCK & CO INC.

PI Jacobson MA;

DR WPI; 95-375695/49.

DR N-PSDB: T00644.

PT Inhibiting prodn. of tumour necrosis factor alpha - with adenosine

PT A2b subtype receptor agonist, useful for treating auto-immune

PT diseases, also method for identifying A2b specific cpds.

PS Disclosure; Fig 1; 96pp; English.

CC Tumour necrosis factor-alpha production, especially in monocytes, is

CC inhibited by contacting the adenosine receptor (especially the A2b

CC subtype) with an adenosine receptor-agonist (adenosine 5'-(N-

CC cyclopropyl)carboxamidoadenosine, 5'-(N-ethyl)carboxamidoadenosine,

CC (R)-N6-phenyl-2-propyladenosine or cyclohexyladenosine. The

CC agonists may be used in the therapy of autoimmune diseases.

CC specifically rheumatoid arthritis, rheumatoid spondylitis,

CC inflammatory bowel disease (ulcerative colitis and Crohns disease),

CC intestinal pathology associated with graft vs. host disease, organ

CC transplant reactions, septic shock, fever and myalgia due

CC to infection and cachexia associated with chronic infections,

CC malignancy and AIDS, pulmonary diseases e.g. sarcoidosis,

CC silicosis, chronic pulmonary inflammatory disease, and adult

CC respiratory distress syndrome.

SQ Sequence 326 AA;

Query Match 17.1%; Score 76; DB 15; Length 326;  
 Best Local Similarity 28.9%; Pred. No. 4.11e+01;  
 Matches 13; Conservative 13; Mismatches 16; Indels 3; Gaps 3;

Db 111 iplyrkmvtprraavaagcwlsvfvgltptmfgwnnlsavera 155

QY 14 FPLTFKSSASPRFLTNVTC-CFA-TVTRIPLSN-KVLTAVDRS 55

Search completed: Wed Dec 22 10:19:28 1999  
 Job time : 18 secs.



(TM)

```

MPsrch_pp  protein - protein database search, using Smith-Waterman algorithm
Run on:      Wed Dec 22 10:29:29 1999;  MasPar time 4.39 Seconds
          -----340.918 Million cell updates/sec -
Tabular output not generated.

```

```

Scoring table:      PAM 150
                   Gap 11

Searched:          77977 seqs, 28268293 residues

Post-processing:    Minimum Match 0%
                   Listing first 45 summaries

```

Statistics: Mean 38.646; Variance 77.834; scale 0.497

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	101	24.3	222	1	YORA_PYRWO	HYPOTHETICAL 24.7 KD P 1.52e+02
2	93	22.4	233	1	ATPS2_CYACA	ATP SYNTHASE A CHAIN P 1.88e+01
3	91	21.9	1574	1	MYOSIN_2	MYOSIN-2 ISOFORM. 3.46e+01
4	89	21.4	544	1	NANB_MYCPI	PHOSPHOMANNOMUTASE (EC 6.32e+01
5	87	20.9	1254	1	MDR3_CABEL	MULTIDRUG RESISTANCE P 1.15e+00
6	87	20.9	1411	1	YK63_CABEL	HYPOTHETICAL 157.6 KD 1.15e+00
7	85	20.4	1226	1	YNAL_CAEEL	PROBABLE INTEGRIN ALPH. 2.06e+00
8	84	20.2	438	1	AMYL_HORVU	ALPHA-AMYLASE TYPE A I 2.75e+00
9	83	20.0	456	1	PUR8_HAETN	ADENYLOSUCCINATE LYASE 3.66e+00
10	83	20.0	830	1	MKT1_YEAST	MKT1 PROTEIN. 3.66e+00
11	82	19.7	203	1	YG12_BACTU	HYPOTHETICAL 22.8 KD P 4.86e+00
12	82	19.7	352	1	YORM_TTVI1	HYPOTHETICAL 38.6 KD P 4.86e+00
13	82	19.7	382	1	RYCF_SHIFL	O-ANTIGEN POLYMERASE. 4.86e+00
14	81	19.5	317	1	MSHR_HUMAN	MELANOCYTE STIMULATING 6.45e+00
15	81	19.5	459	1	ME3L_DROME	PUTATIVE ATP-DEPENDENT 6.45e+00
16	81	19.5	715	1	VGL_SFVIR	CAPSID PROTEIN. 6.45e+00
17	81	19.5	1068	1	YCF0_MARPO	HYPOTHETICAL 127 KD PR 6.45e+00
18	80	19.2	124	1	RFBI_SHIFL	RFBI PROTEIN. 8.53e+00
19	80	19.2	214	1	SC22_YEAST	PROTEIN TRANSPORT PROT 8.53e+00
20	80	19.2	240	1	Y03M_MYCTU	HYPOTHETICAL 25.7 KD P 8.53e+00
21	80	19.2	1095	1	NKC2_MOUSE	BUMETANIDE-SENSITIVE S 8.53e+00
22	80	19.2	1095	1	NKC2_RAT	BUMETANIDE-SENSITIVE S 8.53e+00
23	79	19.0	246	1	CDM_HUMAN	CDM PROTEIN 16C6-AG TU 1.12e+01

RX MEDLINE: 94033298  
 RA KOSTRZEWA M., ZETSCHKE K.;  
 RT "Organization of plastid-encoded ATPase genes and flanking regions  
 including homologues of infB and tsf in the thermophilic red alga  
 RT *Galdieria sulphuraria*.";  
 RL PLANT MOL. BIOL. 23:67-76(1993).  
 CC -!- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL; IT MAY PLAY A  
 CC DIRECT ROLE IN THE TRANSLLOCATION OF PROTONS ACROSS THE MEMBRANE.  
 CC -!- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC  
 CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE  
 CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)  
 CC HAS THREE MAIN SUBUNITS: A, B AND C.  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. CHLOROPLAST  
 CC THYLAKOID MEMBRANE.  
 CC -!- SIMILARITY: BELONGS TO THE ATPASE A CHAIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X67814; G429173;  
 DR FIR: S39515; S39515.  
 DR PROSITE: PS00449; ATPASE\_A; 1.  
 DR PFAM: PF00119; ATP-synt\_A; 1.  
 DR MENDEL: 537; CYACA:atp1.1.  
 KW HYDROGEN ION TRANSPORT; CF(0); CHLOROPLAST; TRANSMEMBRANE; SIGNAL.  
 FT SIGNAL 1 ? POTENTIAL.  
 FT CHAIN ? 233 ATP SYNTHASE A CHAIN.  
 FT TRANSMEM 82 102 POTENTIAL.  
 FT TRANSMEM 121 141 POTENTIAL.  
 FT TRANSMEM 190 210 POTENTIAL.  
 FT TRANSMEM 211 231 POTENTIAL.  
 SQ SEQUENCE 233 AA; 26066 MW; 8407838F CRC32;

Query Match 22.4%; Score 93; DB 1; Length 233;  
 Best Local Similarity 33.3%; Pred. No. 1.89e-01;  
 Matches 11; Conservative 14; Mismatches 7; Indels 1; Gaps 1;

Db 11 HFFFTGLVFKLH-AQVTFVFFVILILFSLV 42  
 QY 6 HLGFKWGVKIFSELYIHGTYDIVLVWTFV 38

RESULT 3  
 ID MYS2\_YEAST STANDARD; PRT; 1574 AA.  
 AC P19524;  
 DT 01-FEB-1991 (REL. 17, CREATED)  
 DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)  
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
 DE MYOSIN-2 ISOFORM.  
 GN MYO2 OR CDC66 OR YOR326W OR O6167.  
 OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).  
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCCHAROMYCETALES;  
 OC SACCCHAROMYCETACEAE; SACCCHAROMYCES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GRF88;  
 RX MEDLINE: 91201404.  
 RA JOHNSTON G.C., PRINDERGAST J.A., SINGER R.A.;  
 RT "The Saccharomyces cerevisiae MYO2 gene encodes an essential myosin  
 RT for vectorial transport of vesicles.";  
 RL J. CELL BIOL. 113:539-551(1991).  
 RN [2]

RP SEQUENCE OF 1-748 FROM N.A.  
 RC STRAIN=S288C / FY1679;  
 RX MEDLINE: 97051589.  
 RA PEARSON B.M., HERNANDO Y., PAYNE J., WOLF S.S., KALOGEROPOULOS A.,  
 RA SCHWEIZER M.;  
 RT "Sequencing of a 35.71 kb DNA segment on the right arm of yeast

RT chromosome XV reveals regions of similarity to chromosomes I and  
 RT XIII.";  
 RL YEAST 12:1021-1031(1996).  
 RN [3]  
 RP SEQUENCE OF 677-1574 FROM N.A.  
 RX MEDLINE: 97051586.  
 RA PARLE-MCDERMOTT A.G., HAND N.J., GOULDING S.G., WOLFE K.H.;  
 RT "Sequence of 29 kb around the PDR10 locus on the right arm of  
 RT Saccharomyces cerevisiae chromosome XV: similarity to part of  
 RT chromosome I.";  
 RL YEAST 12:999-1004(1996).  
 CC -!- FUNCTION: ESSENTIAL MYOSIN FOR VECTORIAL TRANSPORT OF VESICLES.  
 CC PLAYS A DETERMINANT ROLE IN THE SPATIAL REGULATION OF BUD  
 CC FORMATION.  
 CC -!- SUBUNIT: MAY SELF-ASSOCIATE.  
 CC -!- CA(2+) MAY PLAY A ROLE IN MYO2 ACTIVITY.  
 CC -!- SIMILARITY: BELONGS TO CLASS-5 MYOSINS. STRONGEST, TO YEAST MYO4.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: M35532; G172022;  
 DR EMBL: X90565; G940860;  
 DR EMBL: Z49821; E218684;  
 DR EMBL: Z75234; E252153;  
 DR EMBL: Z75235; E332815;  
 DR FIR: A38454; A38454.  
 DR SGD: L0001223; MYO2.  
 DR HSP: PF00063; myosin\_head; 1.  
 DR HSP: P08799; 1MND.  
 KW MYOSIN; COILED COIL; ATP-BINDING; ACTIN-BINDING;  
 KW HEPTAD REPEAT PATTERN; ALKYLATION; MULTIGENE FAMILY; PHOSPHORYLATION;  
 KW CALMODULIN-BINDING.

FT DOMAIN 1 781 GLOBULAR HEAD-LIKE DOMAIN.  
 FT DOMAIN 782 925 NECK.  
 FT DOMAIN 926 981 COILED COIL.  
 FT DOMAIN 982 1009 HINGE.  
 FT DOMAIN 1010 1086 COILED COIL.  
 FT DOMAIN 1087 1574 NON ALPHA-HELICAL, C-TERMINAL DOMAIN.  
 FT DOMAIN 443 523 ACTIN-BINDING (BY SIMILARITY).  
 FT DOMAIN 750 925 ARG/LYS-RICH (BASIC).  
 FT DOMAIN 791 805 CALMODULIN-BINDING (BY SIMILARITY).  
 FT DOMAIN 814 828 CALMODULIN-BINDING (BY SIMILARITY).  
 FT DOMAIN 839 853 CALMODULIN-BINDING (BY SIMILARITY).  
 FT DOMAIN 862 876 CALMODULIN-BINDING (BY SIMILARITY).  
 FT DOMAIN 887 902 CALMODULIN-BINDING (BY SIMILARITY).  
 FT DOMAIN 910 924 CALMODULIN-BINDING (BY SIMILARITY).  
 FT NP\_BIND 164 171 ATP (BY SIMILARITY).  
 FT MOD\_RES 692 692 ALKYLATION (BY SIMILARITY).  
 FT MOD\_RES 702 702 ALKYLATION (BY SIMILARITY).  
 FT MOD\_RES 1097 1097 PHOSPHORYLATION (BY CDC2) (POTENTIAL).  
 SQ SEQUENCE 1574 AA; 180680 MW; AB9D381A CRC32;

\* Query Match 21.9%; Score 91; DB 1; Length 1574;  
 Best Local Similarity 34.4%; Pred. No. 3.46e-01;  
 Matches 11; Conservative 8; Mismatches 12; Indels 1; Gaps 1;

Db 90 IKQVSYOLNIVTYSIV-LIATNPFDRVDQLY 120  
 QY 14 FKIKFSELYIHGTYDIVLVVTFVFSAEAY 45

RESULT 4  
 ID MANB\_MYCPI STANDARD; PRT; 544 AA.  
 AC P47723;  
 DT 01-FEB-1996 (REL. 33, CREATED)  
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

DE PHOSPHOMANNOMUTASE (EC 5.4.2.8) (PM).....  
 GN MANB OR CFSG.  
 OS MYCOPLASMA PIRUM.  
 CC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; MOLLIICUTES;  
 CC MYCOPLASMATACEAE; MYCOPLASMA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BER:  
 RX MEDLINE; 93352438.  
 RA THAM T.N., FERRIS S., KOVACIC R., MONTAGNIER L., BLANCHARD A.;  
 RT "Identification of Mycoplasma pirum genes involved in the salvage  
 RL pathways for nucleosides."  
 CC J. BACTERIOL. 175:5281-5285(1993).  
 CC -!- CATALYTIC ACTIVITY: D-MANNOSE 1-PHOSPHATE - D-MANNOSE 6-PHOSPHATE.  
 CC -!- SIMILARITY: TO OTHER PHOSPHOGLUCOMUTASES AND PHOSPHOMANNOMUTASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; L13289; G401786; ..  
 DR PROSITE; PS00710; PGM\_PMM; 1.  
 DR PFAM; PF00408; PGM\_PMM; 1.  
 KW ISOMERASE; PHOSPHORYLATION.  
 FT ACT\_SITE 145 145  
 FT SEQUENCE 544 AA; 62108 MW; C5FB6254 CRC32;  
 SQ FORMS THE PHOSPHOSERINE INTERMEDIATE  
 (BY SIMILARITY).  
 Query Match 21.4%; Score 89; DB 1; Length 544;  
 Best Local Similarity 36.2%; Pred. No. 6.32e-01;  
 Matches 17; Conservative 11; Mismatches 15; Indels 4; Gaps 3;  
 DB. 362 VYRVGTGFKWGDINKIKDSEFFVVGFEAVGALNSTI-NRDKAY 407  
 QY 2 VFIHLGPKW-G-VFKIKFSELYIHGTDIVLVVFTVFSAEAY 45  
 RESULT 5  
 ID MDR3\_CAEEL STANDARD; PRT; 1254 AA.  
 AC P34713;  
 DT 01-FEB-1994 (REL. 28, CREATED)  
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)  
 DE MULTIDRUG RESISTANCE PROTEIN 3 (P-GLYCOPROTEIN C).  
 GN PGP-3.  
 OS CAENORHABDITIS ELEGANS.  
 CC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEAE; RHABDITIA; RHABDITIDA;  
 CC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE; 93085750.  
 RA LINCKE C.R., THE I., VAN GROENIGEN M., BORST P.;  
 RT "The P-glycoprotein gene family of Caenorhabditis elegans. Cloning  
 RT and characterization of genomic and complementary DNA sequences."  
 RL J. MOL. BIOL. 228:701-711(1992).  
 RN [2]  
 RP TISSUE SPECIFICITY.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE; 93223702.  
 RA LINCKE C.R., BROKS A., THE I., PLASTERK H.A., BORST P.;  
 RT "The expression of two p-glycoprotein (pgp) genes in transgenic  
 RT Caenorhabditis elegans is confined to intestinal cells."  
 RL EMBO J. 12:1615-1620(1993).  
 CC -!- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED  
 CC DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS.  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -!- TISSUE SPECIFICITY: INTESTINAL CELLS.  
 CC -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY

CC (ABC TRANSPORTERS). BELONGS TO THE MDR SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; X65055; G6811; ..  
 DR PIR; S27338; S27338.  
 DR PROSITE; PS60211; ABC\_TRANSPORTER; 2.  
 DR PFAM; PF00005; ABC\_tran; 2.  
 DR PFAM; PF00864; ABC\_membrane; 2.  
 KW ATP-BINDING; GLYCOPROTEIN; TRANSMEMBRANE; TRANSPORT; DUPLICATION;  
 KW MULTIGENE FAMILY.  
 FT DOMAIN 1 46 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 47 67 POTENTIAL.  
 FT TRANSMEM 104 124 POTENTIAL.  
 FT TRANSMEM 173 193 POTENTIAL.  
 FT TRANSMEM 200 220 POTENTIAL.  
 FT TRANSMEM 281 301 POTENTIAL.  
 FT TRANSMEM 310 330 POTENTIAL.  
 FT DOMAIN 331 693 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 694 714 POTENTIAL.  
 FT TRANSMEM 738 758 POTENTIAL.  
 FT TRANSMEM 814 834 POTENTIAL.  
 FT TRANSMEM 835 855 POTENTIAL.  
 FT TRANSMEM 918 938 POTENTIAL.  
 FT TRANSMEM 955 975 POTENTIAL.  
 FT DOMAIN 976 1254 CYTOPLASMIC (POTENTIAL).  
 FT NP\_BIND 411 418 ATP (POTENTIAL).  
 FT NE\_BIND 1050 1057 ATP (POTENTIAL).  
 SO SEQUENCE 1254 AA; 138807 MW; A9AD92D2 CRC32;  
 Query Match 20.9%; Score 87; DB 1; Length 1254;  
 Best Local Similarity 36.4%; Pred. No. 1.15e+00;  
 Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;  
 DB. 698 LVFTLIRGFTWPAFSIVYGOLF 719  
 QY 1 MVFIHLGFKWGVKIKFSELY 22  
 RESULT 6  
 ID YK63\_CAEEL STANDARD; PRT; 1411 AA.  
 AC P34342;  
 DT 01-FEB-1994 (REL. 28, CREATED)  
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE HYPOTHETICAL 157.5 KD PROTEIN C29E4.3 IN CHROMOSOME III.  
 GN C29E4.3.  
 OS CAENORHABDITIS ELEGANS.  
 CC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEAE; RHABDITIA; RHABDITIDA;  
 CC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE; 94150718.  
 RA WILSON R., AINSKOUGH R., ANDERSON K., BAYNES C., BERKS M., COULSON A.,  
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., FRASER A.,  
 RA CRAYTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,  
 RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,  
 RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISIER N.,  
 RA LATREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,  
 RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
 RA SIMS M., SMALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,  
 RA SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,  
 RA WATKESON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,  
 RA WOHLDMAN P.;  
 RT "2.2 Mb cf contiguous nucleotide sequence from chromosome III of C.  
 RT elegans.";

```

EMBL; L23645; G388605; -.
DR DR
DR PIR; S44824; S44824.
DR DR
DR WORMPEP; F54F2.1; CE00194.
DR DR
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR DR
DR PFAM; RF00357; integrin_A; 2.
DR DR

```

DR, PFAM; PF00128; alpha-amylase; 1.  
 DR MENDEL; 8617; HORVU; Amy1; 1.

DR, PFAM; PF00128; alpha-amylase; 1.  
 DR MENDEL; 8617; HORVU; Amy1; 1.

DR HSP; P04063; 1MY.  
KW HYDROLASE; GLYCOSIDASE; CARBOHYDRATE METABOLISM; SEED; GERMINATION;  
KW CALCIUM; MULTIGENE FAMILY; SIGNAL.  
FT SIGNAL 1 24  
FT CHAIN 25 438 ALPHA-AMYLASE TYPE A.  
FT ACT\_SITE 204 204 BY SIMILARITY.  
FT ACT\_SITE 229 229 BY SIMILARITY.  
FT ACT\_SITE 315 315 BY SIMILARITY.  
SQ SEQUENCE 438 AA; 47796 MW; D56DD9FA CRC32;

Query Match 20.2%; Score 84; DB 1; Length 438;  
Best Local Similarity 28.6%; Pred. No. 2.75e+00;  
Matches 14; Conservative 12; Mismatches 20; Indels 3; Gaps 3;

Db 346 IFYDFH-FNMG-FKQIAALVAIRNRGTATSAKILMHEDDAYVAEI 392  
QY 2 VFIHLGFKWGVKIKFSELY-IGHYTDIVLVTVFERSAEAYVHI 49

RESULT 9  
ID PUR8\_HAEN STANDARD; PRT: 456 AA.  
AC P44797;  
DT 01-NOV-1995 (REL. 32, CREATED)  
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
DE ADENYLOSUCGINATE LYASE (EC 4.3.2) (ADENYLOSUCGINASE) (ASL).  
GN PURB OR HI0639.  
OS HAEMOPHILUS INFLUENZAE.  
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURILLACEAE;  
OC HAEMOPHILUS.  
RN  
RP SEQUENCE FROM N.A.  
RC STRAIN-RD / KW20;  
RX MEDLINE: 95350630.  
RA KRELMAN A.R., BULT C.J., TOMS J.F., DOUGHERTY B.A., MERRICK J.M.,  
RA KRELMAN A.R., BULT C.J., TOMS J.F., DOUGHERTY B.A., MERRICK J.M.,  
RA SCOTT J.D., SHIRLEY R., LIU L.I., GLODEK A., KELLEY J.M.,  
RA WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLOM E., COTTON M.D.,  
RA UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,  
RA FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEORGE N.S.M.,  
RA GNEM C.L., McDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,  
RA VENTER J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus  
influenzae Rd."  
RL SCIENCE 269:496-512(1995).  
CC -1- CATALYTIC ACTIVITY: 1-(5-PHOSPHORIBOSYL)-4-(N-SUCCINO-CARBOXAMIDE)  
CC -5-AMINOIMIDAZOLE - FUMARATE + 5'-PHOSPHORIBOSYL-5-AMINO-4-  
CC IMIDAZOLECARBOXAMIDE (ALSO CATALYSES: N6-(1,2-DICARBOXYETHYL)AMP -  
CC FUMARATE + AMP).  
CC -1- PATHWAY: EIGHT STEP IN DE NOVO PURINE BIOSYNTHESIS.  
CC -1- SIMILARITY: TO EUKARYOTIC ADENYLOSUCGINATE AND ARGININOSUCGINATE  
CC LYASES.

This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: U32747; G1573637; -  
TIGR: H10639; -  
DR PROSITE; PS00163; FUMARATE\_LYASES; 1.  
DR PFAM; PF00206; Lyase\_1; 1.  
KW PURINE BIOSYNTHESIS; LYASE.  
SQ SEQUENCE 456 AA; 51993 MW; D545C002 CRC32;

Query Match 20.0%; Score 83; DB 1; Length 456;  
Best Local Similarity 25.5%; Pred. No. 3.56e+00;  
Matches 12; Conservative 12; Mismatches 20; Indels 3; Gaps 3;

Db 233 FVTSIGIQNPYTTQIEPHDYITEFFDVAVYRFTIILDFDLWGYI 279  
QY 3 FTHLGFKGWVKIKF-SELYIGHYTDIVV-LVVFTV-FERSAEAYV 46

RESULT 10  
ID MKT1\_YEAST STANDARD; PRT: 830 AA.  
AC P40850;  
DT 01-FEB-1995 (REL. 31, CREATED)  
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
DE MKT1 PROTEIN.  
GN MKT1 OR YNL085W OR N2302.  
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOCYCETES; SACCHAROMYCETALES;  
OC SACCHAROMYCETACEAE; SACCHAROMYCES.  
RN  
RP SEQUENCE FROM N.A.  
RC STRAIN-2574;  
RX MEDLINE: 95176705.  
RA VERMUT M., WIDNER W.R., DINMAN J.D., WICKNER R.B.;  
RT "Sequence of MKT1, needed for propagation of M2 satellite dsRNA of  
the L-A virus of Saccharomyces cerevisiae."  
RL YEAST 10:1477-1479(1994).  
RN  
RP SEQUENCE FROM N.A.  
RC STRAIN-S288C / FY1679;  
RX MEDLINE: 96310628.  
RA SOLER-MIRA A., SAIZ J.E., BALLESTA J.P.G., REMACHA M.;  
RT "The sequence of a 17,933 bp segment of Saccharomyces cerevisiae  
chromosome XIV contains the RHO2, TOP2, MKT1 and END3 genes and five  
new open reading frames."  
RL YEAST 12:485-491(1996).

-1- FUNCTION: REQUIRED FOR PROPAGATION OF M2 DSRNA SATELLITE AT  
TEMPERATURE ABOVE 30 DEGREES CELSIUS IF AN L-A RNA VIRUS CARRYING  
NEX (N) IS PRESENT IN THE CELLS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: U09129; G520476; -  
DR EMBL; X89016; G887626; -  
DR EMBL; 271361; E239915; -  
DR EMBL; 271350; E328945; -  
DR SGD; L0001120; MKT1.  
DR PFAM; PF00752; XPG\_N; 1.  
FT CONFLICT 30 30 D -> G (IN REF. 1).  
FT CONFLICT 809 830 IDENVEKLETKAVEFTTALSS -> TMKTCNLNHY  
(IN REF. 1).

Query Match 20.08; Score 83; DB 1; Length 830;  
Best Local Similarity 37.9%; Pred. No. 3.66e+00;  
Matches 11; Conservative 11; Mismatches 3; Indels 4; Gaps 4;

Db 595 WGEVFLKLNELNIDSKYHESVIFLNF 622  
QY 10 KWG-VYKIFSELYING-YTD-IVLVVF 35

RESULT 11  
ID YG12\_BACTU STANDARD; PRT: 203 AA.  
AC P10023;  
DT 01-MAR-1989 (REL. 10, CREATED)  
DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)  
DT 01-MAY-1991 (REL. 18, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 22.8 KD PROTEIN (ORF 2).  
OS BACILLUS THURINGIENSIS.

Query Match 20.08; Score 83; DB 1; Length 830;  
Best Local Similarity 37.9%; Pred. No. 3.66e+00;  
Matches 11; Conservative 11; Mismatches 3; Indels 4; Gaps 4;

OG PLASMID PG12.  
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;  
 OC BACILLUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H1.1;  
 RX MEDLINE; 89098342.  
 RA WHILLON J., SEURINCK J.;  
 RT "Complete nucleotide sequence of pGI2, a Bacillus thuringiensis  
 RL plasmid containing Tn430.";  
 RL NUCLEIC ACIDS RES. 16:11827-11827(1988).  
 CC -!- FUNCTION: POSSIBLY INVOLVED IN PG12 REPLICATION MECHANISM.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X13481; G40320; -  
 DR PIR; S02048; S02048.  
 KW HYPOTHETICAL PROTEIN; PLASMID.  
 SQ SEQUENCE 203 AA; 22826 MW; CA914B9A CRC32;  
 Query Match 19.7%; Score 82; DB 1; Length 203;  
 Best Local Similarity 22.5%; Pred. No. 4.86e+00;  
 Matches 9; Conservative 15; Mismatches 15; Indels 1; Gaps 1;  
 DB 110 WGGFTLKAGDIFITNATSSAGIVGHAIA-NGDNYILHMP 148  
 QY 11 WGVFKIKFSELYIHGYTDIVLVVFTVFERSAEAYVHIS 50  
 RESULT 12  
 ID YORM\_TTV1 STANDARD; PRT; 352 AA.  
 AC P19297;  
 DT 01-NOV-1990 (REL. 16, CREATED)  
 DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1990 (REL. 16, LAST ANNOTATION UPDATE)  
 DE HYPOTHETICAL 38.6 KD PROTEIN.  
 OS THERMOPHILEUS TENAX VIRUS 1 (STRAIN KRA1) (TTV1).  
 OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; LIPOTHRIXVIRIDAE;  
 OC LIPOTHRIXVIRUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA NEUMANN H.;  
 RL SUBMITTED (MAR-1989) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X14855; G62178; -  
 KW HYPOTHETICAL PROTEIN  
 SQ SEQUENCE 352 AA; 38674 MW; 51D0560D CRC32;  
 Query Match 19.7%; Score 82; DB 1; Length 352;  
 Best Local Similarity 23.8%; Pred. No. 4.86e+00;  
 Matches 10; Conservative 16; Mismatches 14; Indels 2; Gaps 2;  
 DB 125 LFTVQ-GYPWYNNPNVVFAPIGYTGTFMGIDNVVMMRT 165  
 QY 2 VFIHLGKMGVKEKSELYIH-GYTDIVLVVFTVFERSA 42  
 RESULT 13  
 ID RFC\_SHIFL STANDARD; PRT; 382 AA.

P37784;  
 DT 01-OCT-1994 (REL. 30, CREATED)  
 DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)  
 DE O-ANTIGEN POLYMERASE.  
 GN RFC.  
 OS SHIGELLA FLEXNERI.  
 OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;  
 OC SHIGELLA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SEROTYPE 2A;  
 RX MEDLINE; 94131953.  
 RA MORONA R., MAVRIS M., FALLARINO A., MANNING P.A.;  
 RT "Characterization of the rfc region of Shigella flexneri.";  
 RL J. BACTERIOL. 176:733-747(1994).  
 CC -!- FUNCTION: MAY LINK THE O-ANTIGEN TETRASACCHARIDE UNITS INTO LONG  
 CC CHAINS, GIVING RISE TO TYPICAL SMOOTH LPS.  
 CC -!- PATHWAY: LIPOPOLYSACCHARIDE BIOSYNTHESIS.  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE  
 CC (POTENTIAL).  
 CC -!- SIMILARITY: TO OTHER O-ANTIGEN POLYMERASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X71970; G454905; -  
 KW LIPOPOLYSACCHARIDE BIOSYNTHESIS; TRANSMEMBRANE; INNER MEMBRANE.  
 FT TRANSMEM 11 31 POTENTIAL.  
 FT TRANSMEM 36 56 POTENTIAL.  
 FT TRANSMEM 61 81 POTENTIAL.  
 FT TRANSMEM 87 107 POTENTIAL.  
 FT TRANSMEM 126 146 POTENTIAL.  
 FT TRANSMEM 166 186 POTENTIAL.  
 FT TRANSMEM 194 214 POTENTIAL.  
 FT TRANSMEM 228 248 POTENTIAL.  
 FT TRANSMEM 303 323 POTENTIAL.  
 FT TRANSMEM 325 345 POTENTIAL.  
 FT TRANSMEM 348 368 POTENTIAL.  
 SQ SEQUENCE 382 AA; 43723 MW; 91B8562A CRC32;  
 Query Match 19.7%; Score 82; DB 1; Length 382;  
 Best Local Similarity 27.8%; Pred. No. 4.86e+00;  
 Matches 10; Conservative 12; Mismatches 11; Indels 3; Gaps 3;  
 DB 257 VRFNDLYFY-YKN-VDLATF-LFGRGFGSFLDLRL 289  
 QY 16 IKFSELYIHGYTDIVLVVFTVFERSAEAYVHIS 51  
 RESULT 14  
 ID MSHR\_HUMAN STANDARD; PRT; 317 AA.  
 AC Q01726;  
 DT 01-JUL-1993 (REL. 26, CREATED)  
 DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)  
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
 DE MELANOCYTE STIMULATING HORMONE RECEPTOR (MSH-R) (MELANOTROPIN  
 DE RECEPTOR) (MELANOCORTIN-1 RECEPTOR) (MCL-R).  
 GN MCL1 OR MSHR.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=SKIN.  
 RX MEDLINE; 92390715.  
 RA MOUNTJOY K.G., ROBBINS L.S., MORTUUD M., CONE R.D.;  
 RT "The cloning of a family of genes that encode the melanocortin

receptors.  
SCIENCE 257:1248-1251(1992).  
[2]  
SEQUENCE FROM N.A.  
MEDLINE; 93216807.  
RX GANTZ I., KONDA Y., TASHIRO T., SHIMOTO Y., MIWA H., MUNZERT G.,  
RA WATSON S.J., DELVALLE J., YAMADA T.,  
RT "Molecular cloning of a novel melanocortin receptor."  
RL J. BIOL. CHEM. 268:8246-8250(1993).  
[3]  
SEQUENCE FROM N.A.  
RX MEDLINE; 92387402.  
RA CHAJLANI V., WIKBERG J.E.S.,  
RT "Molecular cloning and expression of the human melanocyte stimulating  
RT hormone receptor cDNA."  
RL FEBS LETT. 309:417-420(1992).  
[4]  
3D-STRUCTURE MODELLING.  
RX MEDLINE; 98304744.  
RA PRUSIS P., SCHIOTH H.B., MUCENIECE R., HERZYK P., AFSHAR M.,  
RA HUBBARD R.E., WIKBERG J.E.,  
RT "Modeling of the three-dimensional structure of the human melanocortin  
RT 1 receptor, using an automated method and docking of a rigid cyclic  
RT melanocyte-stimulating hormone core peptide."  
RL J. MOL. GRAPH. MODEL. 15:307-317(1997).  
[5]  
VARIANTS GLU-84 AND MET-92.  
RX MEDLINE; 97144179.  
RA KOPPULA S.V., ROBBINS L.S., LU D., BAACK E., WHITE C.R. JR.,  
RA SWANSON N.A., CONE R.D.,  
RT "Identification of common polymorphisms in the coding sequence of the  
RT human MSH receptor (MC1R) with possible biological effects."  
RL HUM. MUTAT. 9:30-36(1997).  
CC -!- FUNCTION: RECEPTOR FOR MSH (ALPHA, BETA AND GAMMA) AND ACTH. THE  
CC ACTIVITY OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE  
CC ADENYLATE CYCLASE.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -!- TISSUE SPECIFICITY: MELANOCYTES AND CORTICADRENAL TISSUE.  
CC -!- POLYMORPHISM: VARIATIONS IN MC1R SEEM TO BE LINKED TO THE DEGREE  
CC OF SKIN PIGMENTATION (TYPE I-IV). TYPE I SKIN THE MOST LIGHTLY  
CC PIGMENTED AND TYPE IV THE MOST DARK PIGMENTED.  
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X65634; G34791;  
CC EMBL; X67594; G38410;  
CC PIR; S25558; S25558.  
CC GCRDB; GCR\_0446;  
CC GCRDB; GCR\_0590;  
CC MIM; 155555;  
CC  
CC PROSITE; PS00237; G\_PROTEIN\_RECEPTOR; 1.  
CC PFAM; PF00001; 7tm1; 2.  
KW G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN;  
KW PHOSPHORYLATION; LIPOPROTEIN; PALMITATE; POLYMORPHISM.  
FT DOMAIN 1 37  
FT TRANSMEM 38  
FT DOMAIN 64 72  
FT TRANSMEM 73 93  
FT DOMAIN 94 118  
FT TRANSMEM 119 140  
FT DOMAIN 141 163  
FT TRANSMEM 164 183  
FT DOMAIN 184 191  
FT TRANSMEM 192 211  
FT DOMAIN 212 240  
FT TRANSMEM 241 266

FT DOMAIN 267 279  
FT TRANSMEM 280 300  
FT DOMAIN 301 317  
FT CARBOHYD 29 29  
FT LIPID 315 315  
FT VARIANT 84 84  
FT VARIANT 92 92  
FT CONFLICT 90 90  
FT CONFLICT 162 162  
FT CONFLICT 164 164  
SQ SEQUENCE 317 AA; 34660 MW; 04D89957 CRC32;  
Query Match 19.5%; Score 81; DB 1; Length 317;  
Best Local Similarity 39.3%; Pred. No. 6.45e+00;  
Matches 11; Conservative 5; Mismatches 11; Indels 1; Gaps 1;  
db 169 W VASVVFSTLFAYDYHVAVLCLVWF 195  
QY 11 WSVFVKIKFSELYHGYTDIVLVVFTVF 38  
RESULT 15  
ID H31\_DROME STANDARD; PRT; 459 AA.  
AC P23128;  
DT 01-NOV-1991 (REL. 20, CREATED)  
DF 01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)  
DT 01-NOV-1991 (REL. 20, LAST ANNOTATION UPDATE)  
DE PUTATIVE ATP-DEPENDENT RNA HELICASE.  
GN ME31B;  
OS DROSOPHILA MELANOGASTER (FRUIT FLY).  
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;  
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYROIDEA;  
OC DROSOPHILIDAE; DROSOPHILA.  
RN 14  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 91172763.  
RA DE VALOIR T., TUCKER M.A., BELIKOFF E.J., CAMP L.A., BOLDUC C.,  
RA BECKINGHAM K.,  
RA "A second maternally expressed Drosophila gene encodes a putative RNA  
RA helicase of the 'dead box' family."  
RL PROC. NATL. ACAD. SCI. U.S.A. 88:2213-2217(1991).  
CC -!- FUNCTION: UNWINDS RNA IN AN ATP-DEPENDENT FASHION (POTENTIAL).  
CC -!- DEVELOPMENTAL STAGE: MATERNALLY EXPRESSED (DURING OOGENESIS).  
CC -!- FUNCTION DURING EARLY EMBRYOGENESIS.  
CC -!- SIMILARITY: TO OTHER "DEAD" BOX FAMILY HELICASES. HIGHEST TO  
CC HUMAN P54.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M59925; G157594;  
CC PIR; A39157; A39157.  
CC FLYBASE; Fgtn0004419; me31B.  
CC PROSITE; PS00039; DEAD\_ATP\_HELICASE; 1.  
CC PFAM; PF00270; DEAD; 1.  
CC PFAM; PF00271; helicase\_C; 1.  
KW HELICASE; ATP-BINDING; RNA-BINDING.  
FT NPLBIND 102 109  
FT SITE 207 210  
FT SITE 207 210  
SQ SEQUENCE 459 AA; 52012 MW; 382D89A3 CRC32;  
Query Match 19.5%; Score 81; DB 1; Length 459;  
Best Local Similarity 38.9%; Pred. No. 6.45e+00;  
Matches 14; Conservative 10; Mismatches 9; Indels 3; Gaps 3;  
db 352 SDLFTRG-IDVQAVNVVINFDFPRMAETYLHRIGRS 386  
QY 19 SELZYHGYTDI-VLVVFTVF-FERSNAEYVHISRG 52

Search completed: Wed Dec 22 10:29:37 1999  
Job time : 8 secs.



\*\*\*\*\*  
WILEY  
\*\*\*\*\*  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Mpsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Dec 22 10:29:03 1999; Maspar time 6.42 Seconds  
Tabular output not generated.

Title: >US-09-209-961-21  
Description: (1-53) from US09209961.ppt  
Perfect Score: 416  
Sequence: 1 MVEIHLGFKWGVKIKFSE.....VFTVFSABAYVYVHISRL 53

Scoring table: PAM 150  
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: p1r60  
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 37.675; Variance 87.397; scale 0.431

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	101	24.3	222	1	hypothetical protein	8.59e-02
2	93	22.4	233	2	H+-transporting ATP s	8.02e-01
3	93	22.4	465	2	transhydrogenase chal	8.02e-01
4	91	21.9	1574	1	myosin MYO2 - yeast	1.38e+00
5	89	21.4	544	2	phosphomannomutase (E	2.36e+00
6	87	20.9	276	2	ABC-type transport pr	4.01e+00
7	87	20.9	535	2	cellulase (EC 3.2.1.4	4.01e+00
8	87	20.9	1254	2	P-glycoprotein C - Ca	4.01e+00
9	87	20.9	1411	2	C29E4.3 protein - Cae	4.01e+00
10	86	20.7	434	2	hypothetical protein	5.21e+00
11	85	20.4	315	2	hypothetical protein	6.75e+00
12	85	20.4	1226	2	F54F2.1 protein - Cae	6.75e+00
13	84	20.2	422	2	methicillin resistanc	8.74e+00
14	84	20.2	438	1	alpha-amyrase (EC 3.2	8.74e+00
15	84	20.2	477	2	probable membrane pro	8.74e+00
16	84	20.2	493	2	UMP synthase - fruit	8.74e+00
17	83	20.0	311	2	probable multiple-TM	1.13e+01
18	83	20.0	507	2	adenylosuccinate lyas	1.13e+01
19	83	20.0	830	2	MKT1 protein - yeast	1.13e+01
20	82	19.7	203	2	hypothetical protein	1.46e+01
21	82	19.7	382	2	O-antigen polymerase	1.46e+01
22	82	19.7	383	2	conserved hypothetical	1.46e+01
23	82	19.7	528	2	probable RNA helicase	1.46e+01

24 82 19.7 1224 2 E71611... hypothetical protein- 1.46e+01  
25 62 19.5 131 2 S65197... probable membrane pro 1.87e+01  
26 63 19.5 285 2 A36819... C9 protein - rabbit f 1.87e+01  
27 21 19.5 317 2 S29204... melanotropin receptor 1.87e+01  
28 31 19.5 369 2 A39157... probable RNA helicase 1.87e+01  
29 81 19.5 715 2 H71352... probable sigma factor 1.87e+01  
30 81 19.5 715 2 S08450... hypothetical protein 1.87e+01  
31 81 19.5 787 2 H71453... hypothetical protein 1.87e+01  
32 81 19.5 1068 2 S01519... hypothetical protein 1.87e+01  
33 31 19.5 1148 2 A71446... hypothetical protein 1.87e+01  
34 36 19.2 124 2 D36966... rfc 3'-region hypothe 2.41e+01  
35 80 19.2 152 2 I57438... NADH dehydrogenase (u 2.41e+01  
36 30 19.2 152 2 I77327... NADH dehydrogenase (u 2.41e+01  
37 80 19.2 214 2 S51405... synaptobrevin SEC22 - 2.41e+01  
38 80 19.2 240 2 E70771... hypothetical protein 2.41e+01  
39 80 19.2 245 2 S71116... BCR-associated protei 2.41e+01  
40 70 19.2 411 2 S61933... Ca2+/H+-exchanging pr 2.41e+01  
41 80 19.2 748 2 T02560... cellulose synthase no 2.41e+01  
42 80 19.2 1095 2 A54145... sodium-potassium-cho 2.41e+01  
43 80 19.2 1095 2 I49270... Na+/K+/Cl--cotranspor 2.41e+01  
44 80 19.2 1095 2 I49269... Na+/K+/Cl--cotranspor 2.41e+01  
45 80 19.2 1095 2 I49268... Na+/K+/Cl--cotranspor 2.41e+01

ALIGNMENTS

1  
ENTRY QOQYAW #type complete  
TITLE hypothetical protein A - Pyrococcus woesei  
ORGANISM #formal\_name Pyrococcus woesei  
DATE 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 30-Jun-1993

ACCESSIONS S10652  
REFERENCE S10650  
#authors Zwickl, P.; Fabry, S.; Bogedain, C.; Haas, A.; Hensel, R.  
#journal J. Bacteriol. (1990) 172:4329-4338  
#title Glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic archaeobacterium Pyrococcus woesei: characterization of the enzyme, cloning and sequencing of the gene, and expression in Escherichia coli.

#cross-references MUID:90330536  
#accession S10652  
#molecule\_type DNA  
#residues 1-222 #label ZWI  
SUMMARY #length 222 #molecular-weight 24755 #checksum 5694

Query Match 24.3%; Score 101; DB 1; Length 222;  
Best Local Similarity 32.7%; Pred. No. 8.59e-02;  
Matches 1f; Conservative 12; Mismatches 19; Indels 2; Gaps 2;

Db 113 IISTVAFFSWYSPFWALLITFTGIVGLIVFLVLSKMPSDAIVV 161  
CY 1 MVEIHLGFKWGVKIKFSELYHGTYDIVV-LWVFTVFER-SAEAYV 47

2

ENTRY S39515 #type complete  
TITLE H+-transporting ATP synthase (EC 3.6.1.34) chain a - red alga (Cyanidium caldarium) chloroplast  
ORGANISM #formal\_name Chloroplast Cyanidium caldarium  
DATE 07-Oct-1994 #sequence\_revision 10-Nov-1995 #text\_change 14-Nov-1997

ACCESSIONS S39515  
REFERENCE S39512  
#authors Kostzewa, M.; Zetsche, K.  
#journal Plant Mol. Biol. (1993) 23:67-76  
#title Organization of plastid-encoded ATPase genes and flanking regions including homologues of infB and tsf in the thermophilic red alga Galdieria sulphuraria.

#cross-references MUID:94033298  
#accession S39515  
#status  
#molecule\_type DNA

ACCESSIONS A38454; S58339; S62058; S67232; S71956; S72009

REFERENCE A38454

#authors Johnston, G.C.; Prendergast, J.A.; Singer, R.A.

#journal J. Cell Biol. (1991) 113:539-551

#title The Saccharomyces cerevisiae MYO2 gene encodes an essential myosin for vectorial transport of vesicles.

#cross-references MUID:91201404

#accession A38454

##molecule\_type DNA

##residues 1-1574 #label JOH

##cross-references GB:M35532; NID:g172021; PID:g172022

REFERENCE S58318

#authors Pearson, B.M.; Hernando, Y.; Wolf, S.S.; Kalogeropoulos, A.; Schweizer, M.

#submission submitted to the EMBL Data Library, August 1995

#accession S58339

##molecule\_type DNA

##residues 1-748 #label PEA

##cross-references EMBL:X90565; NID:g940836; PID:g940860

REFERENCE S62058

#authors Parle, A.G.; Hand, N.J.; Goulding, S.G.; Wolfe, K.H.

#submission submitted to the EMBL Data Library, June 1995

#accession S67213

##molecule\_type DNA

##residues 1-748 #label PEW

##cross-references EMBL:Z75234; MIPS:YOR326W

##experimental\_source strain S288C

REFERENCE S67233

#authors Goulding, S.E.; Hand, N.J.; Parle-McDermott, A.G.; Wolfe, K.H.

#submission submitted to the Protein Sequence Database, July 1996

#accession S67232

##molecule\_type DNA

##residues 1-748 #label PEW

##cross-references EMBL:Z75234; MIPS:YOR326W

##experimental\_source strain S288C

REFERENCE S71966

#authors Parle-McDermott, A.G.; Hand, N.J.; Goulding, S.E.; Wolfe, K.H.

#journal Yeast. (1996) 12:999-1004

#title Sequence of 29 kb around the PDR10 locus on the right arm of Saccharomyces cerevisiae chromosome XV: similarity to part of chromosome I.

#cross-references MUID:97051586

#accession S71966

##status nucleic acid sequence not shown; translation not shown

##molecule\_type DNA

##residues 677-1574 #label PAW

##cross-references EMBL:Z49821; NID:g1163062; PID:e218684; PID:g1163063

##note the nucleotide sequence was submitted to the EMBL Data Library, June 1995

REFERENCE S71986

#authors Pearson, B.M.; Hernando, Y.; Payne, J.; Wolf, S.S.; Kalogeropoulos, A.; Schweizer, M.

#journal Yeast. (1996) 12:1021-1031

#title Sequencing of a 35.71 kb DNA segment on the right arm of yeast chromosome XV reveals regions of similarity to chromosomes I and XIII.

#accession S72009

##status nucleic acid sequence not shown; translation not shown

##molecule\_type DNA

##residues 1-748 #label PFF

```

#cross-references EMBL:X90565; NID:g940836; PID:g940860
#note the nucleotide sequence was submitted to the EMBL Data
Library, August 1995
COMMENT The neck domain comprises six approximately 23-residue tandem
repeats; this domain may be responsible for calmodulin binding.
GENETICS
#gene SGD:MYO2
#map_position 15R
#cross-references SGD:S0005853; MIPS:YOR326W
FUNCTION
#description involved in vacuole inheritance; involved in polarized growth
and secretion; plays a role in vesicle transport along
actin cables to the bud site
CLASSIFICATION #superfamily myosin MYO2; myosin motor domain homology
KEYWORDS actin binding; ATP; coiled coil; cytoskeleton; hydrolase;
P-loop; phosphoprotein; tandem repeat
FEATURE
73-769 #domain myosin motor domain homology #label MMOT\
164-171 #region nucleotide-binding motif A (P-loop)\
552-673 #region actin binding #status predicted\
782-926 #domain neck #status predicted #label NEC\
942-1086 #domain coiled coil #status predicted #label COI\
1087-1574 #domain carboxyl-terminal #status predicted #label CTD\
170 #binding_site ATP (Lys) #status predicted\
692,702 #active_site Cys #status predicted\
1097,1452 #binding_site phosphate (Thr) (covalent) #status
predicted
SUMMARY #length 1574 #molecular-weight 180679 #checksum 13
Query Match 21.9%; Score 91; DB 1; Length 1574;
Best Local Similarity 34.4%; Pred. No. 1.38e+00;
Matches 11; Conservative 8; Mismatches 12; Indels 1; Gaps 1;
Db 90 IKRYSQNLNIVYSGIV-LIATNPFDRVDOLY 120
Qy 14 FKRFSELYIHGTDIVLVVFTVFERSAEAY 45

RESULT 5
ENTRY #type complete
TITLE Phosphomannomutase (EC 5.4.2.8) - Mycoplasma pirum (strain
BER) (SGC3)
ORGANISM #formal_name Mycoplasma pirum
DATE 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change
10-Sep-1997
ACCESSIONS E53312
REFERENCE A53312
#authors Blanchard, A.
#journal J. Bacteriol. (1993) 175:5281-5285
#title Identification of Mycoplasma pirum genes involved in the
salvage pathways for nucleosides.
#cross-references MUID:93352438
#accession E53312
#status preliminary
#molecule_type DNA
#residues 1-544 #label THA
#cross-references GB:L13289; NID:g401781; PID:g401786
GENETICS
#genetic_code SGC3
KEYWORDS intramolecular transferase; isomerase
SUMMARY #length 544 #molecular-weight 62108 #checksum 7566
Query Match 21.4%; Score 89; DB 2; Length 544;
Best Local Similarity 36.2%; Pred. No. 2.36e+00;
Matches 17; Conservative 11; Mismatches 15; Indels 4; Gaps 3;
Db 362 VYRGTGFGKWDKINKDKSEFFVGFEEAVGALNSIT-NRDKDAY 407
Qy 2 VFIIHLGFW-G--VFKIKFSELYIHGTDIVLVVFTVFERSAEAY 45

RESULT 6
#cross-references EMBL:X90565; NID:g940836; PID:g940860
#note the nucleotide sequence was submitted to the EMBL Data
Library, August 1995
COMMENT The neck domain comprises six approximately 23-residue tandem
repeats; this domain may be responsible for calmodulin binding.
GENETICS
#gene SGD:MYO2
#map_position 15R
#cross-references SGD:S0005853; MIPS:YOR326W
FUNCTION
#description involved in vacuole inheritance; involved in polarized growth
and secretion; plays a role in vesicle transport along
actin cables to the bud site
CLASSIFICATION #superfamily myosin MYO2; myosin motor domain homology
KEYWORDS actin binding; ATP; coiled coil; cytoskeleton; hydrolase;
P-loop; phosphoprotein; tandem repeat
FEATURE
73-769 #domain myosin motor domain homology #label MMOT\
164-171 #region nucleotide-binding motif A (P-loop)\
552-673 #region actin binding #status predicted\
782-926 #domain neck #status predicted #label NEC\
942-1086 #domain coiled coil #status predicted #label COI\
1087-1574 #domain carboxyl-terminal #status predicted #label CTD\
170 #binding_site ATP (Lys) #status predicted\
692,702 #active_site Cys #status predicted\
1097,1452 #binding_site phosphate (Thr) (covalent) #status
predicted
SUMMARY #length 1574 #molecular-weight 180679 #checksum 13
Query Match 21.9%; Score 91; DB 1; Length 1574;
Best Local Similarity 34.4%; Pred. No. 1.38e+00;
Matches 11; Conservative 8; Mismatches 12; Indels 1; Gaps 1;
Db 90 IKRYSQNLNIVYSGIV-LIATNPFDRVDOLY 120
Qy 14 FKRFSELYIHGTDIVLVVFTVFERSAEAY 45

RESULT 5
ENTRY #type complete
TITLE Phosphomannomutase (EC 5.4.2.8) - Mycoplasma pirum (strain
BER) (SGC3)
ORGANISM #formal_name Mycoplasma pirum
DATE 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change
10-Sep-1997
ACCESSIONS E53312
REFERENCE A53312
#authors Blanchard, A.
#journal J. Bacteriol. (1993) 175:5281-5285
#title Identification of Mycoplasma pirum genes involved in the
salvage pathways for nucleosides.
#cross-references MUID:93352438
#accession E53312
#status preliminary
#molecule_type DNA
#residues 1-544 #label THA
#cross-references GB:L13289; NID:g401781; PID:g401786
GENETICS
#genetic_code SGC3
KEYWORDS intramolecular transferase; isomerase
SUMMARY #length 544 #molecular-weight 62108 #checksum 7566
Query Match 21.4%; Score 89; DB 2; Length 544;
Best Local Similarity 36.2%; Pred. No. 2.36e+00;
Matches 17; Conservative 11; Mismatches 15; Indels 4; Gaps 3;
Db 362 VYRGTGFGKWDKINKDKSEFFVGFEEAVGALNSIT-NRDKDAY 407
Qy 2 VFIIHLGFW-G--VFKIKFSELYIHGTDIVLVVFTVFERSAEAY 45

RESULT 6
#cross-references EMBL:X90565; NID:g940836; PID:g940860
#note the nucleotide sequence was submitted to the EMBL Data
Library, June 1996
COMMENT the nucleotide sequence was submitted to the EMBL Data
Library, June 1996
KEYWORDS transport protein
SUMMARY #length 276 #molecular-weight 31542 #checksum 9527
Query Match 20.9%; Score 87; DB 2; Length 276;
Best Local Similarity 28.9%; Pred. No. 4.01e+00;
Matches 11; Conservative 14; Mismatches 10; Indels 3; Gaps 3;
Db 224 MVSVID-GFRWALIGGE-STIFLPGFL-LSLLELVIIIF 258
Qy 1 MVEIHLGFWGFKIKFSELYIHGTDIVLVVFTVF 38

RESULT 7
ENTRY #type fragment
TITLE cellulase (EC 3.2.1.4) 88K precursor - Prevotella ruminicola
(fragment)
ALTERNATE_NAMES endo-1,4-beta-glucanase
OF_SANISM #formal_name Prevotella ruminicola
DATE 28-May-1992 #sequence_revision 28-May-1992 #text_change
20-Mar-1998
ACCESSIONS A41375; B41375
REFERENCE A41375
#authors Matsushita, O.; Russell, J.B.; Wilson, D.B.
#journal J. Bacteriol. (1991) 173:6919-6926
#title A Bacteroides ruminicola 1,4-beta-D-endoglucanase is encoded
in two reading frames.
#cross-references MUID:92041578
#accession A41375
#status preliminary
#molecule_type DNA
#residues 1-535 #label MAT
#accession B41375
#status preliminary
#molecule_type DNA
#residues 150-535 #label MA2
#cross-references GB:S61753; NID:g238044; PID:g238046
FUNCTION
#description hydrolysis of 1,4-beta-D-glucosidic linkages in
beta-D-glucans such as cellulose and lichenin; can
hydrolyze such linkages in beta-D-glucans that also contain
1,3-linkages
cellulose degradation
#pathway
glycosylase; hydrolase; polysaccharide degradation
KEYWORDS

```

```

ENTRY S74745 #type complete
TITLE ABC-type transport protein slr0977 - Synecocystis sp.
(strain PCC 6803)
ALTERNATE_NAMES protein slr0977
ORGANISM #formal_name Synecocystis sp.
#variety PCC 6803
DATE 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
21-Aug-1998
ACCESSIONS S74745
REFERENCE S74322
#authors Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
Nakamura, Y.; Miyajima, N.; Hirosewa, M.; Sugita, M.;
Saitomoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo,
S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
Yasuda, M.; Tabata, S.
#journal DNA Res. (1996) 3:109-136
#title Sequence analysis of the genome of the unicellular
cyanobacterium Synecocystis sp. PCC6803. II. Sequence
determination of the entire genome and assignment of
potential protein-coding regions.
#cross-references MUID:97061201
#accession S74745
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-276 #label KAN
#cross-references EMBL:D90901; GB:AB001339; NID:g1651897; PID:d1017629;
PID:g1651970
#note the nucleotide sequence was submitted to the EMBL Data
Library, June 1996
KEYWORDS transport protein
SUMMARY #length 276 #molecular-weight 31542 #checksum 9527
Query Match 20.9%; Score 87; DB 2; Length 276;
Best Local Similarity 28.9%; Pred. No. 4.01e+00;
Matches 11; Conservative 14; Mismatches 10; Indels 3; Gaps 3;
Db 224 MVSVID-GFRWALIGGE-STIFLPGFL-LSLLELVIIIF 258
Qy 1 MVEIHLGFWGFKIKFSELYIHGTDIVLVVFTVF 38

RESULT 7
ENTRY #type fragment
TITLE cellulase (EC 3.2.1.4) 88K precursor - Prevotella ruminicola
(fragment)
ALTERNATE_NAMES endo-1,4-beta-glucanase
OF_SANISM #formal_name Prevotella ruminicola
DATE 28-May-1992 #sequence_revision 28-May-1992 #text_change
20-Mar-1998
ACCESSIONS A41375; B41375
REFERENCE A41375
#authors Matsushita, O.; Russell, J.B.; Wilson, D.B.
#journal J. Bacteriol. (1991) 173:6919-6926
#title A Bacteroides ruminicola 1,4-beta-D-endoglucanase is encoded
in two reading frames.
#cross-references MUID:92041578
#accession A41375
#status preliminary
#molecule_type DNA
#residues 1-535 #label MAT
#accession B41375
#status preliminary
#molecule_type DNA
#residues 150-535 #label MA2
#cross-references GB:S61753; NID:g238044; PID:g238046
FUNCTION
#description hydrolysis of 1,4-beta-D-glucosidic linkages in
beta-D-glucans such as cellulose and lichenin; can
hydrolyze such linkages in beta-D-glucans that also contain
1,3-linkages
cellulose degradation
#pathway
glycosylase; hydrolase; polysaccharide degradation
KEYWORDS

```

## SUMMARY #length 535 #checksum 8354

Query Match 20.9%; Score 87; DB 2; Length 535;  
 Best Local Similarity 28.8%; Pred. No. 4.01e+00;  
 Matches 15; Conservative 15; Mismatches 18; Indels 4; Gaps 3;  
 Db 134 VYVINTNGK-SVXXXXXMKFEI--FSLATLISFVFAQSDADYYVYKADG 182  
 QY 2 VFIHLGFKWGVKIFSELYIHGTDIVLVVFTVFSAE-AYVWHISG 52

RESULT 8  
 ENTRY S27338 #type complete  
 TITLE P-glycoprotein C - Caenorhabditis elegans  
 ORGANISM #formal\_name Caenorhabditis elegans  
 DATE 07-May-1993 #sequence\_revision 07-May-1993 #text\_change 10-Jul-1998  
 ACCESSIONS S27338; S19980  
 REFERENCE S27337  
 #authors Lincke, C.R.; The, I.; van Groenigen, M.; Borst, P.  
 #journal J. Mol. Biol. (1992) 228:701-711  
 #title The P-glycoprotein gene family of *Caenorhabditis elegans*.  
 #description Cloning and characterization of genomic and complementary DNA sequences.  
 #accession S27338  
 #status nucleic acid sequence not shown  
 #molecule\_type DNA  
 #residues 1-1254 #label LIN  
 #cross-references EMBL:X65055; NID:g6810; PID:g6811

GENETICS  
 #gene pgp-3  
 #map\_position 10  
 #introns 34/3; 69/1; 116/3; 260/2; 434/3; 502/3; 580/1; 613/1; 689/3;  
 775/3; 957/2; 1170/2  
 CLASSIFICATION  
 #superfamily unassigned ATP-binding cassette proteins;  
 ATP-binding cassette homology  
 ATP; glycoprotein; membrane protein; P-loop  
 FEATURE  
 394-588 #domain ATP-binding cassette homology #label ABC1  
 411-418 #region nucleotide-binding motif A (P-loop)  
 1033-1227 #domain ATP-binding cassette homology #label ABC2  
 1050-1057 #region nucleotide-binding motif A (P-loop)  
 SUMMARY #length 1254 #molecular-weight 138806 #checksum 684

Query Match 20.9%; Score 87; DB 2; Length 1254;  
 Best Local Similarity 36.4%; Pred. No. 4.01e+00;  
 Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;  
 Db 698 LVFTLNGFTWPAFIVYGLF 719  
 QY 1 MVFIHLGFKWGVKIFSELY 22

RESULT 9  
 ENTRY S44770 #type complete  
 TITLE C29E4.3 protein - *Caenorhabditis elegans*  
 ORGANISM #formal\_name *Caenorhabditis elegans*  
 DATE 14-Sep-1994 #sequence\_revision 12-May-1995 #text\_change 09-Sep-1997  
 ACCESSIONS S44770  
 REFERENCE S44737  
 #authors Wilson, R.  
 #description submitted to the EMBL Data Library, September 1993  
 #accession S44770  
 #status preliminary  
 #molecule\_type DNA  
 #residues 1-1411 #label WIL  
 #cross-references EMBL:L23651; NID:g388575; PID:g388581  
 GENETICS  
 #introns 22/1; 115/1; 188/3; 309/2; 363/1; 596/3; 848/1; 882/3; 920/3;  
 945/3; 1058/3; 1119/1; 1161/3; 1181/1; 1290/2; 1323/1  
 SUMMARY #length 1411 #molecular-weight 157562 #checksum 9164

Query Match 20.9%; Score 87; DB 2; Length 1411;  
 Best Local Similarity 34.2%; Pred. No. 4.01e+00;  
 Matches 13; Conservative 11; Mismatches 10; Indels 4; Gaps 4;  
 Db 1093 RFINEYDIIGYEENLV-ISFSAFO-STQKIYITSHVKG 1128  
 QY 17 KFSELY-IHGTDIVLVVFTVFSAEAYV-HISRG 52

RESULT 10  
 ENTRY T01938 #type complete  
 TITLE hypothetical protein F1104.6 - *Arabidopsis thaliana*  
 ORGANISM #formal\_name *Arabidopsis thaliana* #common\_name mouse-ear cress  
 DATE 26-Feb-1999 #sequence\_revision 26-Feb-1999 #text\_change 24-Mar-1999  
 ACCESSIONS T01938  
 REFERENCE Z14466  
 #authors Abu-Threideh, J.; Stoneking, T.; Langston, Y.; Trevaskis, E.  
 #submission submitted to the EMBL Data Library, October 1998  
 #description The sequence of *A. thaliana* F1104.  
 #accession T01938  
 #status translated from GB/EMBL/DBJ  
 #molecule\_type DNA  
 #residues 1-434 #label ABU  
 #cross-references EMBL:AF096370; NID:g3695372; PID:g3695379  
 #experimental\_source cultivar Columbia

GENETICS  
 #map\_position 4  
 #introns 93/3; 120/3; 152/3; 180/3; 204/1; 238/3; 265/2; 298/2; 321/3;  
 364/3  
 #note F1104.6  
 SUMMARY #length 434 #molecular-weight 50274 #checksum 5375

Query Match 20.7%; Score 86; DB 2; Length 434;  
 Best Local Similarity 30.8%; Pred. No. 5.21e+00;  
 Matches 12; Conservative 13; Mismatches 11; Indels 3; Gaps 3;  
 Db 346 FEW-VHKVSPMDTSSFIAEKVLLINSV-FRVSPSFV 382  
 QY 9 PKWGVKIFSELYIHGVTDIVLVVFTVFSAEAYV 46

RESULT 11  
 ENTRY E69365 #type complete  
 TITLE hypothetical protein AF0925 - *Archaeoglobus fulgidus*  
 ORGANISM #formal\_name *Archaeoglobus fulgidus*  
 DATE 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 05-Jun-1998  
 ACCESSIONS E69365  
 REFERENCE A69250  
 #authors Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.L.; Kerlavage, A.R.; Graham, D.E.; Kyrpides, N.C.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeill, L.K.; Badger, J.H.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.; Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.; D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.; Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.

#journal Nature (1997) 390:364-370  
 #title The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon *Archaeoglobus fulgidus*.  
 #cross-references MIM:98049343  
 #accession E69365  
 #status preliminary; nucleic acid sequence not shown; translation not shown  
 #molecule\_type DNA

[illegible]

#submission submitted to the EMBL Data Library, December 1994  
#description The sequence of S. cerevisiae cosmid 8300.  
#accession S51344  
#molecule\_type DNA  
#residues 1-477 #label DUZ  
#cross-references EMBL:U19028; NID:g609380; PID:g609386; MIPS:YLR341w  
GENETICS  
#map\_position 12R  
KEYWORDS transmembrane protein  
FEATURE  
209-225 #domain transmembrane #status predicted #label TM1\  
339-355 #domain transmembrane #status predicted #label TM2  
SUMMARY #length 477 #molecular-weight 55138 #checksum 131  
Query Match 20.2%; Score 84; DB 2; Length 477;  
Best Local Similarity 34.3%; Pred. No. 8.74e+00;  
Matches 12; Conservative 13; Mismatches 7; Indels 3; Gaps 3;  
Db 351 FILIVKMYSVNFEKFSDEVFL-SPIELIKIVF 384  
||:|:|:| ||| |:|:|:|:|:|  
QY 3 FIHLGFKWGV-EKIKFS-ELYIHGYTDIVLVVF 35

Search completed: Wed Dec 22 10:29:13 1999  
Job time : 10 secs.

\*\*\*\*\*  
 W P E R L H  
 (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Dec 22 10:30:26 1999; MasPar time 3.34 Seconds  
 Tabular output not generated. 188,899 Million cell updates/sec

Title: >US-09-209-961-21  
 Description: (1-53) from US09209961.pep  
 Perfect Score: 416  
 Sequence: 1 MVFIHLGFGVGFKIFSE.....VFTVFSRABAYVYHISRGL 53

Scoring table: PAM 150  
 Gap 11

Searched: 122461 seqs, 11912985 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: a-issued  
 1:5A\_COMB 2:5B\_COMB 3:PCT9\_COMB 4:backfiles1

Statistics: Mean 24 697; Variance 110.974; scale 0.223

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	81	19.5	317	1	US-08-672-Sequence 2, Applicatio	1.99e+01
2	81	19.5	317	2	US-08-842-Sequence 2, Applicatio	1.99e+01
3	81	19.5	317	1	US-07-866-Sequence 6, Applicatio	1.99e+01
4	81	19.5	317	1	US-08-671-Sequence 2, Applicatio	1.99e+01
5	81	19.5	317	2	US-08-842-Sequence 2, Applicatio	1.99e+01
6	81	19.5	317	2	US-08-780-Sequence 4, Applicatio	1.99e+01
7	80	19.2	317	2	US-08-466-Sequence 6, Applicatio	2.36e+01
8	79	19.0	404	2	US-09-015-Sequence 2, Applicatio	2.79e+01
9	79	19.0	404	2	US-08-696-Sequence 2, Applicatio	2.79e+01
10	78	18.8	400	1	US-08-118-Sequence 14, Applicatio	3.31e+01
11	78	18.8	400	1	US-08-486-Sequence 14, Applicatio	3.31e+01
12	78	18.8	400	2	US-08-488-Sequence 14, Applicatio	3.31e+01
13	78	18.8	400	2	US-08-474-Sequence 14, Applicatio	3.31e+01
14	77	18.5	584	1	US-08-426-Sequence 36, Applicatio	3.92e+01
15	77	18.5	622	1	US-08-426-Sequence 35, Applicatio	3.92e+01
16	76	18.3	906	1	US-08-220-Sequence 9, Applicatio	4.63e+01
17	76	18.3	906	1	US-08-413-Sequence 9, Applicatio	4.63e+01
18	76	18.3	1289	2	US-08-542-Sequence 2, Applicatio	4.63e+01
19	76	18.3	1289	2	US-08-322-Sequence 2, Applicatio	4.63e+01
20	75	18.0	437	2	US-09-076-Sequence 2, Applicatio	5.47e+01
21	75	18.0	437	2	US-08-538-Sequence 2, Applicatio	5.47e+01
22	74	17.8	453	2	US-08-646-Sequence 27, Applicatio	6.46e+01
23	74	17.8	453	2	US-08-599-Sequence 27, Applicatio	6.46e+01

24 74 17.8 500 1 US-08-117-Sequence 70, Applicatio 6.46e+01  
 25 73 17.5 461 1 US-08-194-Sequence 4, Applicatio 7.63e+01  
 26 72 17.3 357 1 US-08-119-Sequence 36, Applicatio 8.99e+01  
 27 71 17.1 308 1 US-07-828-Sequence 9, Applicatio 1.06e+02  
 28 71 17.1 419 1 US-08-330-Sequence 2, Applicatio 1.06e+02  
 29 71 17.1 420 2 US-08-466-Sequence 2, Applicatio 1.06e+02  
 30 71 17.1 855 4 519526-1 Patent No. 519526 1.06e+02  
 31 70 16.8 432 2 US-08-677-Sequence 8, Applicatio 1.25e+02  
 32 70 16.8 1398 2 US-08-750-Sequence 9, Applicatio 1.25e+02  
 33 69 16.6 194 2 US-08-700-Sequence 23, Applicatio 1.47e+02  
 34 69 16.6 194 2 US-08-700-Sequence 25, Applicatio 1.47e+02  
 35 69 16.6 266 2 US-08-773-Sequence 4, Applicatio 1.47e+02  
 36 69 16.6 280 2 US-08-700-Sequence 15, Applicatio 1.47e+02  
 37 69 16.6 280 2 US-08-700-Sequence 17, Applicatio 1.47e+02  
 38 69 16.6 446 2 US-08-672-Sequence 11, Applicatio 1.47e+02  
 39 69 16.6 462 2 US-08-898-Sequence 2, Applicatio 1.47e+02  
 40 69 16.6 462 2 US-08-898-Sequence 4, Applicatio 1.47e+02  
 41 62 16.6 551 2 US-08-700-Sequence 4, Applicatio 1.47e+02  
 42 69 16.6 737 2 US-08-700-Sequence 21, Applicatio 1.47e+02  
 43 69 16.6 799 2 US-08-700-Sequence 27, Applicatio 1.47e+02  
 44 69 16.6 1876 2 US-08-619-Sequence 2, Applicatio 1.47e+02  
 45 69 16.6 2713 3 PCT-US96-0 Sequence 1, Applicatio 1.47e+02

## ALIGNMENTS

RESULT 1  
 ID US-08-672-109B-2 STANDARD; PRT; 317 AA.

XX XXXXXX

Sequence 2, Application US/08672109B

Sequence 2, Application US/08672109B  
 Patent No. 5710265

GENERAL INFORMATION:

APPLICANT: Yamada, Tadataka

APPLICANT: Gantz, Ira

TITLE OF INVENTION: Genes Encoding Melanocortin Receptors

NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:

ADDRESS: Harness, Dickey & Pierce, P.L.C.

STREET: P.O. Box 828

CITY: Bloomfield Hills

STATE: MI

COUNTRY: US

ZIP: 48302

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/672,109B

FILING DATE: June 27, 1996

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Smith, DeAnn F.

REGISTRATION NUMBER: 36683

REFERENCE/DOCKET NUMBER: 2115-000853DVC

TELEPHONE: (810)641-1600

TELEFAX: (810)641-0270

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 317 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE 317 AA: 34691 MW: 523647 CN;

Query Match 19.5%; Score 81; DB 1; Length 317;  
Best Local Similarity 39.3%; Pred. No. 1.99e+01;  
Matches 11; Conservative 5; Mismatches 11; Indels 1; Gaps 1;

Db 169 W-VASVVFSTLFIAVDHVAVLCLVVF 195  
QY 11 WGVFKIFSELYIHGYTDIVLVVFTVF 38

RESULT 2  
ID US-08-842-045-2 STANDARD; PRT; 317 AA.  
XX  
AC xxxxxx  
XX  
DT  
XX  
DE Sequence 2, Application US/08842045  
XX  
CC Sequence 2, Application US/08842045  
CC Patent No. 5817787  
CC GENERAL INFORMATION:  
CC APPLICANT: Yamada, Tadataka  
CC TITLE OF INVENTION: Genes Encoding Melanocortin Receptors  
CC NUMBER OF SEQUENCES: 23  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Harness, Dickey & Pierce, P.L.C.  
CC STREET: P.O. Box 828  
CC CITY: Bloomfield Hills  
CC STATE: MI  
CC COUNTRY: US  
CC ZIP: 48303  
CC  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC FILING DATE: US/08/842,045  
CC CLASSIFICATION: 536  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Smith, Deann F.  
CC REGISTRATION NUMBER: 36693  
CC REFERENCE/DOCKET NUMBER: 2115-000853DVE  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (810)641-1600  
CC TELEFAX: (810)641-0270  
CC INFORMATION FOR SEQ ID NO: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 317 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 317 AA; 34691 MW; 523647 CN;

Query Match 19.5%; Score 81; DB 2; Length 317;  
Best Local Similarity 39.3%; Pred. No. 1.99e+01;  
Matches 11; Conservative 5; Mismatches 11; Indels 1; Gaps 1;

Db 169 W-VASVVFSTLFIAVDHVAVLCLVVF 195  
QY 11 WGVFKIFSELYIHGYTDIVLVVFTVF 38

RESULT 3  
ID US-07-866-979-6 STANDARD; PRT; 317 AA.  
XX  
AC xxxxxx  
XX  
DT  
XX  
DE Sequence 6, Application US/07866979

XX Sequence 6, Application US/07866979  
CC Patent No. 5532347  
CC GENERAL INFORMATION:  
CC APPLICANT: Cone, Roger D  
CC APPLICANT: Mountjoy, Kathleen G  
CC TITLE OF INVENTION: Melanocyte Stimulating Hormone Receptor  
CC NUMBER OF SEQUENCES: 6  
CC TITLE OF INVENTION: and Uses  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Allegretti & Witcoff, Ltd.  
CC STREET: 10 South Wacker Drive, Suite 3000  
CC CITY: Chicago  
CC STATE: Illinois  
CC COUNTRY: USA  
CC ZIP: 60606  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/07/866,979  
CC FILING DATE: 19920410  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: No. 5532347nan, Kevin E  
CC REGISTRATION NUMBER: 35,303  
CC REFERENCE/DOCKET NUMBER: 92,154  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 312-715-1000  
CC TELEFAX: 312-715-1234  
CC INFORMATION FOR SEQ ID NO: 6:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 317 amino acids  
CC TYPE: AMINO ACID  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 317 AA; 34560 MW; 525101 CN;

Query Match 19.5%; Score 81; DB 1; Length 317;  
Best Local Similarity 39.3%; Pred. No. 1.99e+01;  
Matches 11; Conservative 5; Mismatches 11; Indels 1; Gaps 1;

Db 169 W-VASVVFSTLFIAVDHVAVLCLVVF 195  
QY 11 WGVFKIFSELYIHGYTDIVLVVFTVF 38

RESULT 4  
ID US-08-671-525B-2 STANDARD; PRT; 317 AA.  
XX  
AC xxxxxx  
XX  
DT  
XX  
DE Sequence 2, Application US/08671525B  
XX  
CC Sequence 2, Application US/08671525B  
CC Patent No. 5703220  
CC GENERAL INFORMATION:  
CC APPLICANT: Yamada, Tadataka  
CC APPLICANT: Gantz, Ira  
CC TITLE OF INVENTION: Genes Encoding Melanocortin Receptors  
CC NUMBER OF SEQUENCES: 23  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Harness, Dickey & Pierce, P.L.C.  
CC STREET: P.O. Box 828  
CC CITY: Bloomfield Hills  
CC STATE: MI  
CC COUNTRY: US  
CC ZIP: 48303



CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/671,525B  
CC FILING DATE: June 27, 1996  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Smith, Deann F.  
CC REGISTRATION NUMBER: 36683  
CC REFERENCE/DOCKET NUMBER: 2115-000853DVB  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (810)641-1600  
CC TELEFAX: (810)641-0270  
CC INFORMATION FOR SEQ ID NO: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 317 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 317 AA; 523647 CN;  
SQ

Query Match 19.5%; Score 81; DB 1; Length 317;  
Best Local Similarity 39.3%; Pred. No. 1.99e+01;  
Matches 11; Conservative 5; Mismatches 11; Indels 1; Gaps 1;

Db 169 W-VASVVFSTLFIAVDHVAVLCLVWF 195  
QY 11 WGVFKIRFSELYIHGYTDIVVLVWFVTF 38

RESULT 5  
ID US-08-842-238-2 STANDARD; PRT: 317 AA.  
XX xxxxxx

Sequence 2, Application US/08842238  
Sequence 2, Application US/08842238  
Patent No. 5869257  
GENERAL INFORMATION:  
CC APPLICANT: Yamada, Tadataka  
CC APPLICANT: Gantz, Ira  
CC TITLE OF INVENTION: Genes Encoding Melanocortin Receptors  
CC NUMBER OF SEQUENCES: 23  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Harness, Dickey & Pierce, P.L.C.  
CC STREET: P.O. Box 828  
CC CITY: Bloomfield Hills  
CC STATE: MI  
CC COUNTRY: US  
CC ZIP: 48303  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/842,238  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Smith, Deann F.  
CC REGISTRATION NUMBER: 36683  
CC REFERENCE/DOCKET NUMBER: 2115-000853DVB  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (810)641-1600  
CC TELEFAX: (810)641-0270  
CC INFORMATION FOR SEQ ID NO: 2:

CC --SEQUENCE CHARACTERISTICS:  
CC LENGTH: 317 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 317 AA; 523647 CN;  
SQ

Query Match 19.5%; Score 81; DB 2; Length 317;  
Best Local Similarity 39.3%; Pred. No. 1.99e+01;  
Matches 11; Conservative 5; Mismatches 11; Indels 1; Gaps 1;

Db 169 W-VASVVFSTLFIAVDHVAVLCLVWF 195  
QY 11 WGVFKIRFSELYIHGYTDIVVLVWFVTF 38

RESULT 6  
ID US-08-780-749A-4 STANDARD; PRT: 317 AA.  
XX xxxxxx

Sequence 4, Application US/08780749A  
Sequence 4, Application US/08780749A  
Patent No. 5932779  
GENERAL INFORMATION:  
CC APPLICANT: Lee, Frank  
CC APPLICANT: Huszar, Dennis  
CC APPLICANT: Gu, Wei  
CC TITLE OF INVENTION: SCREENING METHODS FOR COMPOUNDS  
CC USEFUL IN THE REGULATION OF BODY WEIGHT  
CC NUMBER OF SEQUENCES: 10  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Pennie & Edmonds LLP  
CC STREET: 1155 Avenue of the Americas  
CC CITY: New York  
CC STATE: New York  
CC COUNTRY: USA  
CC ZIP: 10036/2711  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Diskette  
CC COMPUTER: IBM Compatible  
CC OPERATING SYSTEM: DOS  
CC SOFTWARE: FastSeq Version 2.0  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/780,749A  
CC FILING DATE: 08-JAN-1997  
CC CLASSIFICATION: 800  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Laura A. Coruzzi  
CC REGISTRATION NUMBER: 30,742  
CC REFERENCE/DOCKET NUMBER: 7853-064  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (212) 790-9090  
CC TELEFAX: (212) 869-8864/9741  
CC TELEX: 66141 PENNIE  
CC INFORMATION FOR SEQ ID NO: 4:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 317 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS:  
CC TOPOLOGY: unknown  
CC MOLECULE TYPE: peptide  
CC SEQUENCE 317 AA; 523647 CN;  
SQ

Query Match 19.5%; Score 81; DB 2; Length 317;  
Best Local Similarity 39.3%; Pred. No. 1.99e+01;  
Matches 11; Conservative 5; Mismatches 11; Indels 1; Gaps 1;

Db 169 W-VASVVFSTLFIAVDHVAVLCLVWF 195  
QY 11 WGVFKIRFSELYIHGYTDIVVLVWFVTF 38

100

## TITLE OF INVENTION: NOVEL HUMAN G-PROTEIN COUPLED RECEPTOR

CC NUMBER OF SEQUENCES: 8  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: SmithKline Beecham Corporation  
CC STREET: 709 Swedeland Road  
CC CITY: King of Prussia  
CC STATE: PA  
CC COUNTRY: USA  
CC ZIP: 19406-2799  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Diskette  
CC COMPUTER: IBM Compatible  
CC OPERATING SYSTEM: DOS  
CC SOFTWARE: FastSeq Version 1.5  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/696,770  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER:  
CC FILING DATE:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Han, William T  
CC REGISTRATION NUMBER: 34,344  
CC REFERENCE/DOCKET NUMBER: TAK50001-2  
CC TELEPHONE: 610-270-5219  
CC TELEFAX: 610-270-5090  
CC TELEX:  
CC INFORMATION FOR SEQ ID NO: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 404 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: peptide  
CC HYPOTHETICAL: NO  
CC ANTI-SENSE: NO  
CC FRAGMENT TYPE: N-terminal  
CC ORIGINAL SOURCE:  
CC SEQUENCE 404 AA; 45323 MW; 927330 CN;  
  
Query Match 19.0%; Score 79; DB 2; Length 404;  
Best Local Similarity 34.4%; Pred. No. 2.79e+01;  
Matches 11; Conservative 9; Mismatches 9; Indels 3; Gaps 3;  
  
Db 121 WGSFSCRTHYFYVMYSSIFFLVCLSV-DR 151  
Qy 11 WGVFKIKFSE-LY-INGYTDIVLVVFTVFER 40  
  
RESULT 10  
ID US-08-118-906-14 STANDARD; PRT; 400 AA.  
XX  
AC xxxxxx  
XX  
DT  
XX  
DE Sequence 14, Application US/08118906  
CC Patent No. 5484590  
CC GENERAL INFORMATION:  
CC APPLICANT: Fukuda, Minoru  
CC TITLE OF INVENTION: Expression of the Developmental I  
CC TITLE OF INVENTION: Antigen By a Cloned Human cDNA Encoding a Member of a  
CC TITLE OF INVENTION: Beta-1,6-N-Acetylglucosaminyltransferase Gene Family  
CC NUMBER OF SEQUENCES: 14  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Campbell and Flores  
CC STREET: 4370 La Jolla Village Drive, Suite 700  
CC CITY: San Diego  
  
Sequence 14, Application US/08118906  
Patent No. 5484590  
GENERAL INFORMATION:  
APPLICANT: Fukuda, Minoru  
TITLE OF INVENTION: Expression of the Developmental I  
TITLE OF INVENTION: Antigen By a Cloned Human cDNA Encoding a Member of a  
TITLE OF INVENTION: Beta-1,6-N-Acetylglucosaminyltransferase Gene Family  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell and Flores  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego

CC STATE: California  
CC COUNTRY: USA  
CC ZIP: 92122  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/118,906  
CC FILING DATE: 09-SEP-1993  
CC CLASSIFICATION: 536  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Campbell, Cathryn A.  
CC REGISTRATION NUMBER: 51,815  
CC REFERENCE/DOCKET NUMBER: P-LJ 9526  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (619) 535-9001  
CC TELEFAX: (619) 535-8949  
CC INFORMATION FOR SEQ ID NO: 14:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 400 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 400 AA; 45854 MW; 855321 CN;  
  
Query Match 18.8%; Score 78; DB 1; Length 400;  
Best Local Similarity 25.0%; Pred. No. 3.31e+01;  
Matches 6; Conservative 14; Mismatches 4; Indels 0; Gaps 0;  
  
Db 3 LFTISVSSVIFIVFVSFNFEGDP 31  
Qy 21 LYINGYTDIVLVVFTVFERSAEA 44  
  
RESULT 11  
ID US-08-485-196-14 STANDARD; PRT; 400 AA.  
XX  
AC xxxxxx  
XX  
DT  
XX  
DE Sequence 14, Application US/08486196  
CC Patent No. 5751420  
CC GENERAL INFORMATION:  
CC APPLICANT: Fukuda, Minoru  
CC TITLE OF INVENTION: Expression of the Developmental I  
CC TITLE OF INVENTION: Antigen By a Cloned Human cDNA Encoding a Member of a  
CC TITLE OF INVENTION: Beta-1,6-N-Acetylglucosaminyltransferase Gene Family  
CC NUMBER OF SEQUENCES: 14  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Campbell and Flores  
CC STREET: 4370 La Jolla Village Drive, Suite 700  
CC CITY: San Diego  
CC STATE: California  
CC COUNTRY: USA  
CC ZIP: 92122  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/486,196  
CC FILING DATE:  
CC CLASSIFICATION: 424  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/118,906  
CC FILING DATE: 09-SEP-1993

```

SQ  SEQUENCE  -400 AA; 45854 MW; 855321 CN;

Query Match      18.8%; Score 78; DB 2; Length 400
Best Local Similarity 25.0%; Pred. No. 3,31e+01;
Matches .6; Conservative 14; Mismatches 4; Indels

Db      8 LFIIISVSSVLIIFVFSVENFNGDP 31
      I I : : : : : I I I I I : : : :
Qy      21 LYIHGYTDIVLVVFTVFERSAEA 44

```

RESULT 14  
ID US-08-426-819A-36 STANDARD: PRT: 584 AA.  
XX  
AC xxxxxx  
XX  
DT  
XX  
DE Sequence 36, Application US/08426819A  
XX  
CC Sequence 36, Application US/08426819A  
CC Patent No. 5723318  
CC GENERAL INFORMATION:  
CC APPLICANT: Yamaguchi, No. 5723318omi  
CC APPLICANT: Kojima, Tetsuo  
CC APPLICANT: Oh-Eda, Masayoshi  
CC APPLICANT: Hattori, Kunihiro  
CC TITLE OF INVENTION: Genes Coding for Megakaryocyte  
CC TITLE OF INVENTION: Potentiator  
CC NUMBER OF SEQUENCES: 37  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Birch, Stewart, Kolasch & Birch  
CC STREET: P.O. Box 747  
CC CITY: Falls Church  
CC STATE: Virginia  
CC COUNTRY: USA  
CC ZIP: 22040-0747  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent In Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/426.819A  
CC FILING DATE: 21-APR-1995  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Murphy Jr., Gerald M.  
CC REGISTRATION NUMBER: 28,977  
CC REFERENCE/DOCKET NUMBER: 230-107P  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 703-205-8000  
CC TELEFAX: 703-205-8050  
CC INFORMATION FOR SEQ ID NO: 36:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 584 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: not relevant  
CC TOPOLOGY: not relevant  
CC MOLECULE TYPE: protein  
CC HYPOTHETICAL: NO  
CC FRAGMENT TYPE: internal  
CC SEQUENCE 584 AA; 64098 MW; 1744236 CN;  
Query Match 18.5%; Score 77; DB 1; Length 584;  
Best Local Similarity 47.4%; Pred. No. 3.92e+01;  
Matches 9; Conservative 5; Mismatches 5; Indels 0; Gaps 0;  
Db 313 VLKHKLDLYPQGYPESVI 331  
Qy 13 VFRIKFSELYIHGYTDIVV 31

RESULT 15  
ID US-08-426-819A-35 STANDARD: PRT: 622 AA.  
XX  
AC xxxxxx  
XX  
DT  
XX  
DE Sequence 35, Application US/08426819A  
XX  
CC Sequence 35, Application US/08426819A  
CC Patent No. 5723318

CC GENERAL INFORMATION:  
CC APPLICANT: Yamaguchi, No. 5723318omi  
CC APPLICANT: Kojima, Tetsuo  
CC APPLICANT: Oh-Eda, Masayoshi  
CC APPLICANT: Hattori, Kunihiro  
CC TITLE OF INVENTION: Genes Coding for Megakaryocyte  
CC TITLE OF INVENTION: Potentiator  
CC NUMBER OF SEQUENCES: 37  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Birch, Stewart, Kolasch & Birch  
CC STREET: P.O. Box 747  
CC CITY: Falls Church  
CC STATE: Virginia  
CC COUNTRY: USA  
CC ZIP: 22040-0747  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent In Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/426.819A  
CC FILING DATE: 21-APR-1995  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Murphy Jr., Gerald M.  
CC REGISTRATION NUMBER: 28,977  
CC REFERENCE/DOCKET NUMBER: 230-107P  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 703-205-8000  
CC TELEFAX: 703-205-8050  
CC INFORMATION FOR SEQ ID NO: 35:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 622 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 622 AA; 68037 MW; 1973385 CN;  
Query Match 18.5%; Score 77; DB 1; Length 622;  
Best Local Similarity 47.4%; Pred. No. 3.92e+01;  
Matches 9; Conservative 5; Mismatches 5; Indels 0; Gaps 0;  
Db 351 VLKHKLDLYPQGYPESVI 369  
Qy 13 VFRIKFSELYIHGYTDIVV 31

Search completed: Wed Dec 22 10:30:33 1999  
Job time : 7 secs.

\*\*\*\*\*  
W P S R L H  
\*\*\*\*\*  
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Wed Dec 22 10:14:55 1999; MasPar time 37.19 Seconds  
649.311 Million cell updates/sec  
Tabular output not generated.

Title: >US-09-209-961-3  
Description: (1-314) from US09209961.pep  
Perfect Score: 2358  
Sequence: 1 MPSKNGRSGPQPKRWFT.....GGQFVTLSPPCFPPEYINY 314

Scoring table: PAM 150  
Gap 11  
Searched: 541047 seqs, 76911114 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-pending  
1: P9 2: 060 3: 07 4: 080 5: 081 6: 082 7: 083 8: 084A 9: 084B  
10: 085 11: 086 12: 087 13: 088 14: 089 15: 090 16: 091 17: 092  
18: 093 19: NEWP 20: NEWU6 21: NEWU8 22: NEWU9  
Statistics: Mean 36.699; Variance 134.433; scale 0.273

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES							
Result No.	Score	Match	Length	ID	Description	Pred. No.	
1	2358	100.0	314	22	US-09-209-Sequence 3, Applicatio	8.53e-255	
2	1988	84.3	312	22	US-09-209-Sequence 4, Applicatio	1.35e-211	
3	1988	84.3	312	17	US-09-267-Sequence 21, Applicati	1.35e-211	
4	1988	84.3	312	1	PCT-US99-0Sequence 21, Applicati	1.35e-211	
5	648	27.5	220	1	PCT-US99-0Sequence 8, Applicatio	4.19e-57	
6	648	27.5	220	17	US-09-267-Sequence 8, Applicatio	4.19e-57	
7	197	8.4	286	9	US-08-484-Sequence 16, Applicati	3.05e-08	
8	193	8.2	285	9	US-08-484-Sequence 17, Applicati	7.57e-08	
9	139	5.9	286	9	US-08-484-Sequence 18, Applicati	1.07e-02	
10	98	4.2	561	16	US-09-107-Sequence 5323, Applic	3.77e-01	
11	97	4.1	340	18	US-09-328-Sequence 5334, Applic	4.54e-01	
12	96	4.1	411	16	US-09-107-Sequence 4942, Applic	5.45e-01	
13	97	4.1	611	1	PCT-US96-0Sequence 4, Applicatio	4.54e-01	
14	97	4.1	511	11	US-08-600-Sequence 4, Applicatio	4.54e-01	
15	94	4.0	120	20	US-60-162-Sequence 574, Applicat	7.86e-01	
16	92	3.9	284	12	US-08-751-Sequence 8, Applicatio	1.13e-02	
17	91	3.9	409	20	US-60-164-Sequence 3363, Applic	1.13e-02	
18	92	3.9	479	20	US-60-164-Sequence 66, Applicati	1.13e-02	
19	92	3.9	971	2	US-60-096-Sequence 19531, Applic	1.13e-02	
20	92	3.9	971	17	US-09-248-Sequence 19531, Applic	1.13e-02	
21	92	3.9	2227	9	US-08-475-Sequence 2, Applicatio	1.13e+02	

22	52	3.9	2227	7	US-08-397-Sequence 2, Applicatio	1.13e+02
23	52	3.9	2227	21	US-08-475-Sequence 6, Applicatio	1.13e+02
24	52	3.9	2227	7	US-08-397-Sequence 2, Applicatio	1.13e+02
25	52	3.9	2227	7	US-08-397-Sequence 4, Applicatio	1.13e+02
26	92	3.9	2227	21	US-08-475-Sequence 2, Applicatio	1.13e+02
27	85	3.8	332	16	US-09-134-Sequence 4723, Applic	1.93e+02
28	90	3.8	353	20	US-60-164-Sequence 12463, Applic	1.62e+02
29	90	3.8	400	20	US-60-164-Sequence 12370, Applic	1.62e+02
30	90	3.8	420	15	US-09-079-Sequence 223, Applicat	1.62e+02
31	89	3.8	456	10	US-08-566-Sequence 2, Applicatio	1.93e+02
32	89	3.8	456	10	US-08-566-Sequence 2, Applicatio	1.93e+02
33	89	3.8	456	10	US-08-566-Sequence 2, Applicatio	1.93e+02
34	89	3.8	456	10	US-08-566-Sequence 2, Applicatio	1.93e+02
35	90	3.8	510	16	US-09-134-Sequence 5141, Applic	1.62e+02
36	90	3.8	808	20	US-60-167-Sequence 18734, Applic	1.93e+02
37	90	3.8	1056	15	US-09-079-Sequence 217, Applicat	1.62e+02
38	90	3.8	1130	10	US-08-519-Sequence 6, Applicatio	1.62e+02
39	90	3.8	1437	20	US-60-164-Sequence 4032, Applic	1.93e+02
40	90	3.8	1463	12	US-08-747-Sequence 3, Applicatio	1.62e+02
41	90	3.8	2227	21	US-08-475-Sequence 4, Applicatio	1.62e+02
42	90	3.8	2227	9	US-08-475-Sequence 4, Applicatio	1.62e+02
43	90	3.8	4536	15	US-09-079-Sequence 1, Applicatio	1.62e+02
44	90	3.8	4563	16	US-09-108-Sequence 1, Applicatio	1.62e+02
45	90	3.8	4563	16	US-09-108-Sequence 1, Applicatio	1.62e+02

ALIGNMENTS

RESULT 1  
ID US-09-209-961-3 STANDARD: PRT: 314 AA.  
XX XXXXXX

DE Sequence 3, Application US/09209961

XX Sequence 3, Application US/09209961

CC GENERAL INFORMATION:

CC APPLICANT: WANG, LI

CC APPLICANT: BABIOK, LORNE A.

CC APPLICANT: POTTER, ANDREW A.

CC APPLICANT: WILLSON, PHILIP

CC TITLE OF INVENTION: POSTWEANING MULTISYSTEM WASTING SYNDROME VIRUS FROM

CC TITLE OF INVENTION: PIGS

CC FILE REFERENCE: 9000-0040

CC CURRENT APPLICATION NUMBER: US/09/209,961

CC CURRENT FILING DATE: 1998-12-10

CC EARLIER APPLICATION NUMBER: 60/069,233

CC EARLIER FILING DATE: 1997-12-11

CC EARLIER APPLICATION NUMBER: 60/069,750

CC EARLIER FILING DATE: 1997-12-16

CC NUMBER OF SEQ ID NOS: 24

CC SEQ ID NOS

CC TYPE: PAT

CC JRG/MSM: Porcine Circovirus Type II

CC SEQUENCE 314 AA; 35793 MW; 558374 CN;

Query Match: 100.0%; Score 2358; DB 22; Length 314;  
Best Loca: Similarity 100.0%; Pred. No. 8.53e-255;  
Matches 314; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 MPSKNGRSGPQPKRWFTLNPNSEDERKKIRELPSLFDYFIVGEEGNEGRTPHLOQ 60

1 MPSKNGRSGPQPKRWFTLNPNSEDERKKIRELPSLFDYFIVGEEGNEGRTPHLOQ 60

Db 61 FANFVKQTFNFKVYLGARCHIEKAGTDQONKEYCSKGNLLIECGAPRSGQSRDLS 120

61 FANFVKQTFNFKVYLGARCHIEKAGTDQONKEYCSKGNLLIECGAPRSGQSRDLS 120

Db 121 TAVSTLLESGILTVAEQHPVTFVKNGRGLAELLKLLYSGRMKQKRDWKTNVHFIIVGPPCCGK 180

121 TAVSTLESILVTAQHPVTFKNGRLAELLKVSQKMDKNTNHFIVGPPCGK 180  
181 SKWAANFANPETTYKPPKKNWGDYHGEVWVDDFYGLPWDDLRLCDRYPLTVKTK 240  
181 SKWAANFANPETTYKPPKKNWGDYHGEVWVDDFYGLPWDDLRLCDRYPLTVKTK 240  
241 GGTVPFLARSILITSNOTPLEWYSSSTAVPAVEALYRRTITSLVFWKNATKOSTEGGOFVT 300  
241 GGTVPFLARSILITSNOTPLEWYSSSTAVPAVEALYRRTITSLVFWKNATKOSTEGGOFVT 300  
301 LSPPCPEFPEYIN 314  
301 LSPPCPEFPEYIN 314

RESULT 2  
ID US-09-209-961-4 STANDARD; PRT; 312 AA.

XX xxxxxx

Sequence 4, Application US/09209961

Sequence 4, Application US/09209961

GENERAL INFORMATION:

APPLICANT: WANG, LI

APPLICANT: BABIUK, LORNE A.

APPLICANT: POTTER, ANDREW A.

APPLICANT: WILLSON, PHILIP

TITLE OF INVENTION: POSTWEANING MULTISYSTEM WASTING SYNDROME VIRUS FROM

FILE REFERENCE: 9000-0040

CURRENT APPLICATION NUMBER: US/09/209,961

CURRENT FILING DATE: 1998-12-10

EARLIER APPLICATION NUMBER: 60/069,233

EARLIER FILING DATE: 1997-12-11

EARLIER APPLICATION NUMBER: 60/069,750

EARLIER FILING DATE: 1997-12-16

NUMBER OF SEQ ID NOS: 24

SOFTWARE: Patent In Ver. 2.0

SEQ ID NO 4

LENGTH: 312

TYPE: PRT

ORGANISM: Porcine Circovirus Type I

SEQUENCE 312 AA; 35704 MW; 527440 CN;

Query Match 84.3%; Score 1988; DB 22; Length 312;

Best Local Similarity 84.1%; Pred. No. 1.35e-211;

Matches 259; Conservative 22; Mismatches 26; Indels 1; Gaps 1;

Db 5 KSGPQPKRWVFTLNNPSEEEKKIRELPISLFDYFVCGEEGLEEGRTPHLQGFANFAK 64

QY 8 RSGPQPKRWVFTLNNPSEEDERKIRELPISLFDYFVCGEEGLEEGRTPHLQGFANFAK 67

Db 65 QTFNKVWYFGARCHIEKAKGTDOONKEYCSKEGHIIECGAPNQRSDLSSTAVSTLL 124

QY 68 QTFNKVWYLGARCHIEKAKGTDOONKEYCSKEGNLLIECGAPRSQQRSDLSSTAVSTLL 127

Db 125 ETGSLVTVAEQHPVTFVKNFRGLAELLKVSQKMDKNTNHFIVGPPCGKSQWAAFN 184

QY 128 ESGILVTVAEQHPVTFVKNFRGLAELLKVSQKMDKNTNHFIVGPPCGKSQWAAFN 187

Db 185 AEPRDTYKPSRNKWDYHGEVWVDDFYGLPWDDLRLCDRYPLTVETKGTVPFL 244

QY 188 ANPETTYKPPKKNWGDYHGEVWVDDFYGLPWDDLRLCDRYPLTVETKGTVPFL 247

Db 245 ARSLITSNQAPQEWYSSSTAVPAVEALYRRTITLQFWKTAGEQSTVEPEGRFEAVDPCCA 304

QY 248 ARSLITSNQPLEWYSSSTAVPAVEALYRRTITSLVFWKNATKOSTE-EGGOFVTLSPPCP 306

Db 305 LFPYKINY 312

QY 307 EFPYKINY 314

RESULT 4

ID PCT-US99-05485-21 STANDARD; PRT; 312 AA.

XX xxxxxx

Sequence 21, Application PC/TUS9905485

Sequence 21, Application PC/TUS9905485

GENERAL INFORMATION:

-305 LFPYKINY 312

CY 307 EFPYKINY 314

RESULT 3

ID US-09-267-177-21 STANDARD; PRT; 312 AA.

XX xxxxxx

Sequence 21, Application US/09267177

Sequence 21, Application US/09267177

GENERAL INFORMATION:

APPLICANT: Poet, Steven E.

APPLICANT: Ritchie, Branson W.

APPLICANT: Niagro, Frank D.

APPLICANT: Lukert, Phil D.

TITLE OF INVENTION: Vaccines against Circovirus Infections

FILE REFERENCE: 21099.0057

CURRENT APPLICATION NUMBER: US/09/267,177

CURRENT FILING DATE: 1999-03-12

EARLIER APPLICATION NUMBER: 60/077,890

EARLIER FILING DATE: 1998-03-13

NUMBER OF SEQ ID NOS: 41

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 21

LENGTH: 312

TYPE: PRT

ORGANISM: porcine circovirus

SEQUENCE 312 AA; 35704 MW; 527440 CN;

Query Match 84.3%; Score 1988; DB 17; Length 312;

Best Local Similarity 84.1%; Pred. No. 1.35e-211;

Matches 259; Conservative 22; Mismatches 26; Indels 1; Gaps 1;

Db 5 KSGPQPKRWVFTLNNPSEEEKKIRELPISLFDYFVCGEEGLEEGRTPHLQGFANFAK 64

QY 8 RSGPQPKRWVFTLNNPSEEDERKIRELPISLFDYFVCGEEGLEEGRTPHLQGFANFAK 67

Db 65 QTFNKVWYFGARCHIEKAKGTDOONKEYCSKEGHIIECGAPNQRSDLSSTAVSTLL 124

QY 68 QTFNKVWYLGARCHIEKAKGTDOONKEYCSKEGNLLIECGAPRSQQRSDLSSTAVSTLL 127

Db 125 ETGSLVTVAEQHPVTFVKNFRGLAELLKVSQKMDKNTNHFIVGPPCGKSQWAAFN 184

QY 128 ESGILVTVAEQHPVTFVKNFRGLAELLKVSQKMDKNTNHFIVGPPCGKSQWAAFN 187

Db 185 AEPRDTYKPSRNKWDYHGEVWVDDFYGLPWDDLRLCDRYPLTVETKGTVPFL 244

QY 188 ANPETTYKPPKKNWGDYHGEVWVDDFYGLPWDDLRLCDRYPLTVETKGTVPFL 247

Db 245 ARSLITSNQAPQEWYSSSTAVPAVEALYRRTITLQFWKTAGEQSTVEPEGRFEAVDPCCA 304

QY 248 ARSLITSNQPLEWYSSSTAVPAVEALYRRTITSLVFWKNATKOSTE-EGGOFVTLSPPCP 306

Db 305 LFPYKINY 312

QY 307 EFPYKINY 314

RESULT 4

ID PCT-US99-05485-21 STANDARD; PRT; 312 AA.

XX xxxxxx

Sequence 21, Application PC/TUS9905485

Sequence 21, Application PC/TUS9905485

GENERAL INFORMATION:

-305 LFPYKINY 312





RESULT 7  
ID US-08-484-735-16 STANDARD; PRT; 286 AA.

XXXXXX

Sequence 16, Application US/08484735

Sequence 16, Application US/08484735

GENERAL INFORMATION:

APPLICANT: Wu, Rey-Yuh

APPLICANT: You, Li-Ru

APPLICANT: Soong, Tai-Seng

TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA

TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS AND METHOD FOR DET

TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS

NUMBER OF SEQUENCES: 18

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Neave

STREET: 1251 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: United States

ZIP: 10020

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Word Perfect 5.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/484,735

FILING DATE: 07-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/418,071

FILING DATE: 06-APR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Haley Jr., James F.

REGISTRATION NUMBER: 27,794

REFERENCE/DOCKET NUMBER: DCB-1 CONT

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 596-9000

TELEFAX: (212) 596-9090

INFORMATION FOR SEQ ID NO: 16:

SEQUENCE CHARACTERISTICS:

LENGTH: 286 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

DESCRIPTION: /desc="DNA I-V1"

SEQUENCE 286 AA; 33108 MW; 434824 CN;

Query Match 8.4%; Score 197; DB 9; Length 286;

Best Local Similarity 36.5%; Pred. No. 3.05e-08;

Matches 42; Conservative 21; Mismatches 47; Indels 5; Gaps 5;

Db 4 PSLKWCFTLNYSSAAERENFLSLKEEDVHVYVGVDEVPATGKHLGGYLSLKKRIRLG 63

QY 13 PHKRWFTLNNPSEDERKKIRELPISL-FDYFVGEENEGRTPHLOGFANFYKQTFN 71

Db 64 GLAKKY-GSRAHWEIARGTDEENSKYCK-GTLLLELGFVPVNGSNRKKISEMVA 116

QY 72 KVKW-YLGARCHIEKAGTDQONKEYCSKEGNLLIECGAPRSQG-QRSDLSSTVS 124

RESULT 8

ID US-08-484-735-17 STANDARD; PRT; 285 AA.

XXXXXX

Sequence 18, Application US/08484735

Sequence 18, Application US/08484735

GENERAL INFORMATION:

APPLICANT: Wu, Rey-Yuh

APPLICANT: You, Li-Ru

XX

DE

XX

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC APPLICANT: Soong, Tai-Seng  
CC TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA  
CC TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS AND METHOD FOR DETECTION  
CC TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS  
CC NUMBER OF SEQUENCES: 18  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Fish & Neave  
CC STREET: 1251 Avenue of the Americas  
CC CITY: New York  
CC STATE: New York  
CC COUNTRY: United States  
CC ZIP: 10020  
CC  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Word Perfect 5.0  
CC  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/484,735  
CC FILING DATE: 07-JUN-1995  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/418,071  
CC FILING DATE: 06-APR-1995  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Haley Jr., James F.  
CC REGISTRATION NUMBER: 27,794  
CC REFERENCE/DOCKET NUMBER: DCB-1 CONT  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (212) 596-9000  
CC TELEFAX: (212) 596-9090  
CC INFORMATION FOR SEQ ID NO: 18:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 286 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC DESCRIPTION: /desc="DNA II V2 C1"  
CC SEQUENCE 286 AA; 33482 MW; 434044 CN;  
CC  
CC Query Match 5.9%; Score 139; DB 9; Length 286;  
CC Best Local Similarity 29.8%; Pred. No. 1.07e-02;  
CC Matches 25; Conservative 28; Mismatches 23; Indels 8; Gaps 8;  
CC  
CC Db 8 WMEFINPT-TLPV-MRD-EIK-YKVQV-DRG-OEG-TRHVGVENKRRSSLKQMRGF 60  
CC Y 17 WFTLNPSDEKRIKRLPSLDYFIVGEGNEGRTPHLOGFANFVKKOTFNKVKWY 76  
CC  
CC Db 61 FPGH-HLEKRGSGOEARSCYCKE 83  
CC Y 77 LGARCHIEKAGTDQONKEYCSKE 100  
CC  
CC RESULT 10  
CC ID US-09-107-5323 STANDARD; PRT; 561 AA.  
CC XX  
CC AC xxxxxx  
CC XX  
CC DT  
CC DE  
CC DE Sequence 5323, Application US/09107532  
CC CC  
CC GENERAL INFORMATION:  
CC APPLICANT: Lynn A. Doucette-Stamm and David Bush  
CC TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
CC TITLE OF INVENTION: ENTEROCOCCUS FAECIUM FOR  
CC TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
CC NUMBER OF SEQUENCES: 7308  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
CC STREET: 100 Beaver Street  
CC CITY: Waltham

CC STATE: Massachusetts  
CC COUNTRY: USA  
CC ZIP: 02354  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: CD-ROM ISO9660  
CC COMPUTER:  
CC OPERATING SYSTEM:  
CC SOFTWARE:  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/09/107,532  
CC FILING DATE:  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 60/ 085598  
CC FILING DATE: May 14, 1998  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 60/051571  
CC FILING DATE: July 2, 1997  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Ariniello, Pamela Deneke  
CC REGISTRATION NUMBER: 40,489  
CC REFERENCE/DOCKET NUMBER: GTC-012  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (781)893-5007  
CC TELEFAX: (781)893-8277  
CC INFORMATION FOR SEQ ID NO: 5323:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 561 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC HYPOTHETICAL: YES  
CC ORIGINAL SOURCE:  
CC ORGANISM: Enterococcus faecium  
CC FEATURE:  
CC NAME/KEY: misc\_feature  
CC LOCATION: 1...561  
CC SEQUENCE 561 AA; 67422 MW; 1648944 CN;  
CC  
CC Query Match 4.2%; Score 98; DB 16; Length 561;  
CC Best Local Similarity 28.0%; Pred. No. 3.77e+01;  
CC Matches 14; Conservative 14; Mismatches 20; Indels 2; Gaps 2;  
CC  
CC Db 485 LLEEGILTFRNEYQLIFPTDYGREVLLISTSKVHRKPWADLDYFIV 534  
CC Y 126 LLESGILTVADQHPVTFV-KNERGLA-ELLKVSGRKQKRDWNTNVHFV 173  
CC  
CC RESULT 11  
CC ID US-09-328-3523 STANDARD; PRT; 340 AA.  
CC XX  
CC AC xxxxxx  
CC XX  
CC DT  
CC DE  
CC DE Sequence 5334, Application US/09328352  
CC CC  
CC GENERAL INFORMATION:  
CC APPLICANT: Gary L. Breton et al.  
CC TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOB  
CC TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
CC FILE REFERENCE: GTC99-03PA  
CC CURRENT APPLICATION NUMBER: US/09/328,352  
CC CURRENT FILING DATE: 1999-06-04  
CC NUMBER OF SEQ ID NOS: 8252  
CC SEQ ID NO 5334  
CC LENGTH: 340  
CC TYPE: PRT  
CC ORGANISM: Acinetobacter baumannii  
CC SEQUENCE 340 AA; 37869 MW; 571301 CN;  
CC  
CC Query Match 4.1%; Score 97; DB 18; Length 340;  
CC Best Local Similarity 28.4%; Pred. No. 4.54e+01;





\*\*\*\*\*  
WATERMAN  
\*\*\*\*\*  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Wed Dec 22 10:13:15 1999; MasPar time 24.23 Seconds  
707.243 Million cell updates/sec  
Tabular output not generated.

Title: >US-09-209-961-3  
Description: (1-314) from US09209961.pep  
Perfect Score: 2358  
Sequence: 1 MPKNGRSGPQPHKRWFT.....GGQFVTLSPCPPEFFPEINY 314

Scoring table: PAM 150  
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: sptrmb19

1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phase 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 46.363; Variance 75.952; scale 0.610

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	2351	99.7	314	14	P35.8.	0.00e+00
2	2296	97.4	314	14	PUTATIVE REP PROTEIN (	0.00e+00
3	2293	97.2	314	14	PUTATIVE REP PROTEIN.	0.00e+00
4	1988	84.3	312	14	PUTATIVE PCV REPLICATI	0.00e+00
5	1988	84.3	312	14	REPLICATION-ASSOCIATED	0.00e+00
6	820	34.8	299	14	REPLICATION-ASSOCIATED	1.32e-167
7	242	10.3	290	14	ORF 1.	8.77e-29
8	239	10.1	280	14	POSSIBLE REPLICATION A	3.91e-28
9	224	9.5	282	14	COMPONENT 1-EG.	6.40e-25
10	221	9.4	284	14	ORF V1.	2.77e-24
11	203	8.6	285	14	(BBTV DNA II).	1.65e-20
12	196	8.3	286	14	(BBTV DNA I) V1 AND C1	4.58e-19
13	136	5.8	286	14	ORF V1 (COMPONENT 1).	1.96e-07
14	136	5.8	286	14	HYPOTHETICAL 33.5 KD P	1.96e-07
15	130	5.5	308	14	POLYPROTEIN.	2.28e-06
16	120	5.1	144	14	REPLICATION-ASSOCIATED	1.20e-04
17	119	5.0	148	14	REPLICATION-ASSOCIATED	1.76e-04
18	119	5.0	278	14	FENYV REPLICASE.	3.63e-03
19	111	4.7	102	14	COMPONENT 2.	3.63e-03
20	112	4.7	135	14	REPLICASE B (FRAGMENT)	2.50e-03

21	107	4.5	153	14	O57304	1.57e-02
22	107	4.5	153	14	O83478	1.57e-02
23	107	4.5	153	14	O73572	1.57e-02
24	106	4.5	153	14	O73477	2.25e-02
25	102	4.3	147	14	O86645	9.28e-02
26	102	4.3	148	14	O67567	9.28e-02
27	102	4.3	281	14	O91252	9.28e-02
28	102	4.3	230	10	O23376	9.28e-02
29	99	4.2	286	14	O39828	2.62e-01
30	97	4.1	244	11	O70278	5.17e-01
31	96	4.1	273	10	O43381	7.23e-01
32	97	4.1	275	8	O32904	5.17e-01
33	97	4.1	611	2	O96155	5.17e-01
34	97	4.1	741	2	O60081	5.17e-01
35	96	4.1	1777	14	O89278	7.23e-01
36	95	4.0	94	2	P96226	1.01e+00
37	95	4.0	286	10	O40839	1.01e+00
38	94	4.0	298	10	O40839	1.40e+00
39	95	4.0	435	2	O06721	1.01e+00
40	94	4.0	825	5	O21430	1.40e+00
41	94	4.0	946	2	P70888	1.40e+00
42	92	3.9	183	14	O57160	2.69e+00
43	92	3.9	490	5	O26855	2.69e+00
44	92	3.9	746	14	O67813	2.69e+00
45	92	3.9	2218	14	O67817	2.69e+00

ALIGNMENTS

RESULT 1  
ID O92285 PRELIMINARY; PRT; 314 AA.  
AC O92285;  
DI 01-NOV-1998 (TREMBLREL. 08, CREATED)  
FT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE P35.8.  
OS PORCINE CIRCOVIRUS.  
OC VIRUSES; SSDNA VIRUSES; CIRCOVIRIDAE; CIRCOVIRUS.  
EA [1]  
FP SEQUENCE FROM N.A.  
PC STRAIN:412;  
RA WANG J., WILLSON P., CHOW B., GIBBONS E., BABIUK L.;  
RT "Emergence of a new porcine circovirus."  
RL SUBMITTED (AUG-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: AF085695; G3668363;  
SQ SEQUENCE 314 AA; 35792 MW; F8319091 CRC32;

Query Match 99.7%; Score 2351; DB 14; Length 314;  
Pair Local Similarity 99.7%; Pred. No. 0.00e+00;  
Matches 313; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

60	1	MPKNGRSGPQPHKRWFTLNPNPSEDERKKIRLPISLFDYFIVGEGNEGTRPHLQ	60
61	1	MPKNGRSGPQPHKRWFTLNPNPSEDERKKIRLPISLFDYFIVGEGNEGTRPHLQ	60
62	61	FANFVKQTFNKVYLGARCHEKAGTDOQNKVCSEKGNLLIECGAPRSGQRSDLS	120
63	61	FANFVKQTFNKVYLGARCHEKAGTDOQNKVCSEKGNLLIECGAPRSGQRSDLS	120
64	121	TAVSTLLESSILTVAKQHPVTFVKNFRLGLAELLKVSQKMKRDKNTNVHFI	180
65	121	TAVSTLLESSILTVAKQHPVTFVKNFRLGLAELLKVSQKMKRDKNTNVHFI	180
66	181	SKWAANFANPETTYWPKPKKWDGTHGKRVVDDFYGLWPLWDDLRLCDRPL	240
67	181	SKWAANFANPETTYWPKPKKWDGTHGKRVVDDFYGLWPLWDDLRLCDRPL	240
68	241	GGVTPPLARSILTSNOTPLEWSSSTAVPAVEALYRRTSLVFWKNAKSTEGG	300
69	241	GGVTPPLARSILTSNOTPLEWSSSTAVPAVEALYRRTSLVFWKNAKSTEGG	300
70	301	LSPPC7EFFPEINY 314	

QY 301 LSPPCPEPPEINY 314

## RESULT 2

ID O93192 PRELIMINARY: PRT: 314 AA.  
 AC O93192;  
 DT 01-NOV-1998 (TREMBREL. 08, CREATED)  
 DT 01-NOV-1998 (TREMBREL. 08, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBREL. 08, LAST ANNOTATION UPDATE)  
 DE PUTATIVE REP PROTEIN (OREF1).  
 OS PORCINE CIRCOVIRUS.  
 OC VIRUSES; SSNA VIRUSES; CIRCOVIRIDAE; CIRCOVIRUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-PORCINE CIRCOVIRUS TYPE II;  
 RX MEDLINE; 98418498.  
 RA MEEHAN B.M., MCNEILLY F.M., TODD D., KENNEDY S., JEWURST V.,  
 RA ELLIS J.A., HASSARD L.E., CLARK E.G., HAINES D.M., ALLAN G.M.;  
 RT "Characterisation of novel circovirus DNAs associated with wasting  
 syndromes in pigs.";  
 RL J. GEN. VIROL. 79:2171-2199(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-PORCINE CIRCOVIRUS TYPE II;  
 RA MEEHAN B.M.;  
 RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL; AF055394; G3598832;  
 DR EMBL; AF055393; G3598821;  
 SQ SEQUENCE 314 AA; 35782 MW; 1C88736B CRC32;

Query Match 97.48; Score 2296; DB 14; Length 314;

Best Local Similarity 96.48; Pred. No. 0.00e+00;

Matches 304; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Db 1 MPKKNRSGPQPHKRWFTLNPNSEDERKKIRDLPSLFDYFVIGEGNEGRTPHLQ 60  
 QY 1 MPKKNRSGPQPHKRWFTLNPNSEDERKKIRDLPSLFDYFVIGEGNEGRTPHLQ 60  
 Db 61 FANFVKQTFNKVKYLGARHIEKAGTDOONKEYCSKEGNLMECGAPRSQGRSDLS 120  
 QY 61 FANFVKQTFNKVKYLGARHIEKAGTDOONKEYCSKEGNLMECGAPRSQGRSDLS 120  
 Db 121 TAVSTLLESGSLTVAEQHPVTFVNFPRGLAELLKVSCKMOKRDWKTNNVHVIVGPPCGK 180  
 QY 121 TAVSTLLESGSLTVAEQHPVTFVNFPRGLAELLKVSCKMOKRDWKTNNVHVIVGPPCGK 180  
 Db 181 SKWAANFADPETTYKPPKRWGDIYHGEVVDYFVGLPWDDLLRLCDRPLTVTK 240  
 QY 181 SKWAANFADPETTYKPPKRWGDIYHGEVVDYFVGLPWDDLLRLCDRPLTVTK 240  
 Db 241 GGTVPFLARSILITSNOTPLEWYSSAVPAVEALYRRITSLVFWKNATEQSTEGGFVT 300  
 QY 241 GGTVPFLARSILITSNOTPLEWYSSAVPAVEALYRRITSLVFWKNATEQSTEGGFVT 300  
 Db 301 LSPPCPEPPEINY 314  
 QY 301 LSPPCPEPPEINY 314

## RESULT 3

ID O56123 PRELIMINARY: PRT: 314 AA.  
 AC O56123;  
 DT 01-JUN-1998 (TREMBREL. 06, CREATED)  
 DT 01-JUN-1998 (TREMBREL. 06, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBREL. 08, LAST ANNOTATION UPDATE)  
 DE PUTATIVE REP PROTEIN.  
 OS PORCINE CIRCOVIRUS.  
 OC VIRUSES; SSNA VIRUSES; CIRCOVIRIDAE; CIRCOVIRUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-PWNS PCV;  
 RX MEDLINE; 98241772.  
 RA HAMEL A.L., LIN L.L., NAYAR G.P.S.;

\*Nucleotide sequence of porcine circovirus associated with  
 postweaning multisystemic wasting syndrome in pigs.\*;  
 RL J. VIROL. 72:5262-5267(1998).  
 RN [2].

RC STRAIN-PWNS PCV;  
 RA HAMEL A.L., LIN L.L., NAYAR G.P.S.;  
 RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL; AF027217; G2689646;  
 SQ SEQUENCE 314 AA; 35804 MW; 228DA1CE CRC32;

Query Match 97.28; Score 2293; DB 14; Length 314;

Best Local Similarity 97.18; Pred. No. 0.00e+00;

Matches 305; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Db 1 MPKKNRSGPQPHKRWFTLNPNSEDERKKIRDLPSLFDYFVIGEGNEGRTPHLQ 60  
 QY 1 MPKKNRSGPQPHKRWFTLNPNSEDERKKIRDLPSLFDYFVIGEGNEGRTPHLQ 60  
 Db 61 FANFVKQTFNKVKYLGARHIEKAGTDOONKEYCSKEGNLMECGAPRSQGRSDLS 120  
 QY 61 FANFVKQTFNKVKYLGARHIEKAGTDOONKEYCSKEGNLMECGAPRSQGRSDLS 120  
 Db 121 TAVSTLLESGSLTVAEQHPVTFVNFPRGLAELLKVSCKMOKRDWKTNNVHVIVGPPCGK 180  
 QY 121 TAVSTLLESGSLTVAEQHPVTFVNFPRGLAELLKVSCKMOKRDWKTNNVHVIVGPPCGK 180  
 Db 181 SKWAANFADPETTYKPPKRWGDIYHGEVVDYFVGLPWDDLLRLCDRPLTVTK 240  
 QY 181 SKWAANFADPETTYKPPKRWGDIYHGEVVDYFVGLPWDDLLRLCDRPLTVTK 240  
 Db 241 GGTVPFLARSILITSNOTPLEWYSSAVPAVEALYRRITSLVFWKNATEQSTEGGFVT 300  
 QY 241 GGTVPFLARSILITSNOTPLEWYSSAVPAVEALYRRITSLVFWKNATEQSTEGGFVT 300  
 Db 301 LSPPCPEPPEINY 314  
 QY 301 LSPPCPEPPEINY 314

## RESULT 4

ID P89032 PRELIMINARY: PRT: 312 AA.  
 AC P89032;  
 DT 01-MAY-1997 (TREMBREL. 03, CREATED)  
 DT 01-MAY-1997 (TREMBREL. 03, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBREL. 08, LAST ANNOTATION UPDATE)  
 DE PUTATIVE PCV REPLICATION-ASSOCIATED PROTEIN (REP).  
 OS PORCINE CIRCOVIRUS.  
 OC VIRUSES; SSNA VIRUSES; CIRCOVIRIDAE; CIRCOVIRUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-KIDNEY;  
 RX MEDLINE; 97163487.  
 RA MEEHAN B.M., CREELAN J.L., MCNUITY S.M., TODD D.;  
 RT "Sequence of porcine circovirus DNA: affinities with plant  
 circoviruses.";  
 RL J. GEN. VIROL. 78:221-227(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-KIDNEY;  
 RA MEEHAN B.M.;  
 RL SUBMITTED (FEB-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL; U49186; G1841515;  
 DR PFAM; PF00799; Geminivirus\_all; 1.  
 SQ SEQUENCE 312 AA; 35704 MW; 743D427B CRC32;

Query Match 84.38; Score 1988; DB 14; Length 312;

Best Local Similarity 84.18; Pred. No. 0.00e+00;

Matches 259; Conservative 22; Mismatches 26; Indels 1; Gaps 1;

Db 5 KSGPQPHKRWFTLNPNSEDERKKIRDLPSLFDYFVIGEGNEGRTPHLQGFANFAK 64  
 QY 8 RSGPQPHKRWFTLNPNSEDERKKIRDLPSLFDYFVIGEGNEGRTPHLQGFANFAK 67

```

Db 55 QTFNKVYFGARHIEKAGTDQONKEYCSKEGHLIECGAPRQKGRSDLSSTVSTLL 124
Qy 58 QTFNKVYTLGARCHIEKAGTDQONKEYCSKEGHLIECGAPRQKGRSDLSSTVSTLL 127
Db 125 ETGSLVTVAEQFPVTVYRNFRGLAELLKVGSKQKQDQKTAHVIVGPGCGKSKQWAFNF 184
Qy 128 ESGILVTVAEQHPVTVYRNFRGLAELLKVGSKQKQDQKTAHVIVGPGCGKSKQWAFNF 187
Db 185 APRDTYKPKSRNKNWGDYHGEVVVLDYFGWLPWDDLRLCDRLPLVETKGGTVPEL 244
Qy 188 ANPETTKPKPKNKNWGDYHGEVVVLDYFGWLPWDDLRLCDRLPLVETKGGTVPEL 247
Db 245 ARSILITSNQAPQEWYSSSTAVPAVEALYRITITLQFWKTAGSQSTVEPGRFEAVDPPCA 304
Qy 248 ARSILITSNQPLEWYSSSTAVPAVEALYRITITLQFWKTAGSQSTVEPGRFEAVDPPCA 306
Db 305 LPPYKINY 312
Qy 307 EPPYEINY 314

RESULT 5
ID O90238 PRELIMINARY; PRT; 312 AA.
AC O90238
DT 01-NOV-1998 (TREMREL. 08, CREATED)
DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE REPLICATION-ASSOCIATED PROTEIN.
GN V1.
OS PORCINE CIRCOVIRUS.
OC VIRUSES; SSNA VIRUSES; CIRCOVIRIDAE; CIRCOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 99003925.
RA NIAGRO F.D., FORSTHOEFEL A.N., LAWTHOR R.P., KAMALANATHAN L.,
RA RITCHIE B.W., LATIMER K.S., LUKERT P.D.;
RT "Beak and feather disease virus and porcine circovirus genomes:
RT intermediates between the geminiviruses and plant circoviruses."
RL ARCH. VIROL. 143:1723-1744 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RA NIAGRO F.D., FORSTHOEFEL A.N., LAWTHOR R.P., KAMALANATHAN L.,
RA RITCHIE B.W., LATIMER K.S., LUKERT P.D.;
RL SUBMITTED (JUN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AF071879; G3551844;
FT NON-TER 1
FT SEQUENCE 312 AA; 35665 MW; 8138CD07 CRC32;
SQ SEQUENCE 312 AA; 35665 MW; 8138CD07 CRC32;

Query Match 84.3%; Score 1988; DB:14; Length 312;
Best Local Similarity 83.8%; Pred.No. 0.00e+00;
Matches 258; Conservative 23; Mismatches 26; Indels 1; Gaps 1;

Db 5 KSGPQPHKRWVTLNPNSEERKNKRELPSLDFYVCGEGLEBGRTHLQGFANFAK 64
Qy 8 RSGPQPHKRWVTLNPNSEERKNKRELPSLDFYVCGEGLEBGRTHLQGFANFAK 67
Db 65 QTFNKVYFGARHIEKAGTDQONKEYCSKEGHLIECGAPRQKGRSDLSSTVSTLL 124
Qy 68 QTFNKVYTLGARCHIEKAGTDQONKEYCSKEGHLIECGAPRQKGRSDLSSTVSTLL 127
Db 125 ETGSLVTVAEQFPVTVYRNFRGLAELLKVGSKQKQDQKTAHVIVGPGCGKSKQWAFNF 184
Qy 128 ESGILVTVAEQHPVTVYRNFRGLAELLKVGSKQKQDQKTAHVIVGPGCGKSKQWAFNF 187
Db 185 APRDTYKPKSRNKNWGDYHGEVVVLDYFGWLPWDDLRLCDRLPLVETKGGTVPEL 244
Qy 188 ANPETTKPKPKNKNWGDYHGEVVVLDYFGWLPWDDLRLCDRLPLVETKGGTVPEL 247
Db 245 ARSILITSNQAPQEWYSSSTAVPAVEALYRITITLQFWKTAGSQSTVEPGRFEAVDPPCA 304
Qy 248 ARSILITSNQPLEWYSSSTAVPAVEALYRITITLQFWKTAGSQSTVEPGRFEAVDPPCA 306
Db 305 LPPYKINY 312
Qy 307 EPPYEINY 314

```

```

Qy 307 EPPYEINY 314

RESULT 6
ID O90235 PRELIMINARY; PRT; 299 AA.
AC O90235
DT 01-NOV-1998 (TREMREL. 08, CREATED)
DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE REPLICATION-ASSOCIATED PROTEIN (FRAGMENT).
GN V1.
OS BEAK AND FEATHER DISEASE VIRUS.
OC VIRUSES; SSNA VIRUSES; CIRCOVIRIDAE; CIRCOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 99003925.
RA NIAGRO F.D., FORSTHOEFEL A.N., LAWTHOR R.P., KAMALANATHAN L.,
RA RITCHIE B.W., LATIMER K.S., LUKERT P.D.;
RT "Beak and feather disease virus and porcine circovirus genomes:
RT intermediates between the geminiviruses and plant circoviruses."
RL ARCH. VIROL. 143:1723-1744 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RA NIAGRO F.D., FORSTHOEFEL A.N., LAWTHOR R.P., KAMALANATHAN L.,
RA RITCHIE B.W., LATIMER K.S., LUKERT P.D.;
RL SUBMITTED (JUN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AF071878; G3551840;
FT NON-TER 1
FT SEQUENCE 299 AA; 34479 MW; 248AAC02 CRC32;
SQ SEQUENCE 299 AA; 34479 MW; 248AAC02 CRC32;

Query Match 34.8%; Score 820; DB:14; Length 299;
Best Local Similarity 45.5%; Pred.No. 1.32e-167;
Matches 125; Conservative 50; Mismatches 90; Indels 10; Gaps 9;

Db 20 RWCFTLNPNPTDGEIEFVRSLSGDPDEYFVYALVGRKEBQD-TPHLQGFHFHFKRLSALK 78
Qy 15 KRWFVTLNPNSEDERKKRELPSLDFYFVCGEGLEBGRTHLQGFANFAKQTFNKVK 74
Db 79 KML-PRGHPERAKGSDADNEKCYSGEGDVLTLGIIVARDGHR-EDGAVAVMSPKMK 136
Qy 75 WYLGARHIEKAGTDQONKEYCSKEGHLIECGAPRQKGRSDLSSTVSTLLSGLVLT 134
Db 137 VAREPFDIVYRHRGRLSHLSLLVSGSRP-RDFKTEVDVYIYVPGCGKSRWA-N-EQPGTKY 193
Qy 135 VAEQHPVTVYRNFRGLAELLKVGSKQKQDQKTAHVIVGPGCGKSKQWAFANPETTY 194
Db 194 YKM-RGEMWDYDGEDVILDDYFGLPYCEMLRLCDRLPLVETKGGTVPEL 252
Qy 195 WRPPKKNWGDYHGEVVVLDYFGWLPWDDLRLCDRLPLVETKGGTVPELARSILIT 254
Db 253 SNKAPETWTKEDCDP-KPLFRFRFTR-VWYVNIIDK 284
Qy 255 SNOTPLEWYSSSTAVPAVEALYRITITLQFWKTAGSQSTVEPGRFEAVDPPCA 289

RESULT 7
ID Q66005 PRELIMINARY; PRT; 290 AA.
AC Q66005
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE ORF 1.
OS COCONUT FOLIAR DECAY VIRUS.
OC VIRUSES; SSNA VIRUSES; CIRCOVIRIDAE.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 90266484.
RA KORDE W., RANDLES J.W., LANGRIDGE P., HANOLD D.;
RT "Nucleotide sequence of a circular single-stranded DNA associated
RT with coconut foliar decay virus."
RL VIROLOGY 176:648-651 (1990).
DR EMBL: M29963; G323307;

```

SQ SEQUENCE 290 AA; 33497 MW; 87A78804 CRC32;  
Query Match 10.3%; Score 242; DB 14; Length 290;  
Best Local Similarity 38.2%; Pred. No. 8.77e-29;  
Matches 42; Conservative 26; Mismatches 36; Indels 6; Gaps 6;  
Db 6 RRCFTLYNETEAEANVVRISLNLV-YAIVGDEVAPSTGQRHLOGFTHLTKGRLOG 64  
15 KRWFLLNPNSEDERKKI-REL-PISLFDYFVIGEGNEEGRTPHLOGFANFVKQTFNKV 72  
Db 65 LKTVIGNDRHLPTGSDQNDYCSKE-RVLEHGVPTRPGVKRPLA 113  
QY 73 VKWYLGAR-CHIEKAKGTDOONKEYCSKEGNLLIEGAPRSQG-QRSDLS 120  
RESULT 8  
ID Q87009 PRELIMINARY; PRT; 280 AA.  
AC Q87009;  
DT 01-NOV-1996 (TREMBLREL.. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL.. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL.. 08, LAST ANNOTATION UPDATE)  
DE POSSIBLE REPLICATION ASSOCIATED PROTEIN (SCSV2).  
GN SCSV2.  
OS SUBTERRANEAN CLOVER STUNT VIRUS  
OC VIRUSES; SSDNA VIRUSES; CIRCOVIRIDAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 95193233.  
RA BOEVINK P., CHU P.W., KESE P.;  
RT "Sequence of subterranean clover stunt virus DNA: affinities with the  
RT geminiviruses";  
RL Virology 207:354-361(1995).  
DR EMBL; U16731; G571488;  
SQ SEQUENCE 280 AA; 32473 MW; 3A60B6C7 CRC32;  
Query Match 10.18%; Score 239; DB 14; Length 280;  
Best Local Similarity 41.2%; Pred. No. 3.91e-28;  
Matches 40; Conservative 20; Mismatches 34; Indels 3; Gaps 3;  
Db 3 RRYCFTLNTATEIERETFLSFDLSEDELNYFVGVGDEATTGOK-HLOGFVGFKNKIRLGL 61  
QY 15 KRWFLLNPNSEDERKKI-REL-PISLFDYFVIGEGNEEGRTPHLOGFANFVKQTFNKV 73  
Db 62 KKGFGNRAHWEIARGSDQNRDYCKE-TLISEIGIP 97  
QY 74 KWLIGARCHIEKAKGTDOONKEYCSKEGNLLIEGAP 110  
RESULT 9  
ID Q91254 PRELIMINARY; PRT; 282 AA.  
AC Q91254;  
DT 01-NOV-1998 (TREMBLREL.. 08, CREATED)  
DT 01-NOV-1998 (TREMBLREL.. 08, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL.. 08, LAST ANNOTATION UPDATE)  
DE COMPONENT 1-EG.  
GN C1-EG.  
OS FABA BEAN NECROTIC YELLOW VIRUS.  
OC VIRUSES; SSDNA VIRUSES; CIRCOVIRIDAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-EVI-93;  
RA KATUL L., TIMCHENKO T., GRONENBORN B., VETTEN H.J.;  
RT "The faba bean necrotic yellows virus genome comprises at least ten  
RT circular ssDNA components four of which encode putative Rep  
RT proteins";  
RL SUBMITTED (MAY-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; AJ005968; E1318733;  
SQ SEQUENCE 282 AA; 32676 MW; 9D32A16E CRC32;  
Query Match 9.58%; Score 224; DB 14; Length 282;  
Best Local Similarity 39.6%; Pred. No. 6.40e-25;  
Matches 40; Conservative 18; Mismatches 41; Indels 2; Gaps 2;

Db 4 KRWCFTLYNKTAVPRESFISLFSRDELNYFVCGDETATTNOKHLOGYVSLKKMIRLGL 63  
QY 15 KRWFLLNPNSEDERKKI-REL-PISLFDYFVIGEGNEEGRTPHLOGFANFVKQTFNKV 73  
Db 64 KKGFGNRAHWEIARGSDQNRDYCKE-TLISEIGAPVK 103  
QY 74 KWLIGARCHIEKAKGTDOONKEYCSKEGNLLIEGAPRSQG 114  
RESULT 10  
ID Q89685 PRELIMINARY; PRT; 284 AA.  
AC Q89685;  
DT 01-NOV-1996 (TREMBLREL.. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL.. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1996 (TREMBLREL.. 01, LAST ANNOTATION UPDATE)  
DE ORF VI.  
OS BANANA BUNCHY TOP VIRUS.  
OC VIRUSES; SSDNA VIRUSES; CIRCOVIRIDAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX WU R.Y., YOU L.R.;  
RT SUBMITTED (FEB-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
RL EMBL; U12587; G642394;  
DR EMBL; U12586; G642391;  
SQ SEQUENCE 284 AA; 33174 MW; 7E694C83 CRC32;  
Query Match 9.4%; Score 221; DB 14; Length 284;  
Best Local Similarity 39.3%; Pred. No. 2.77e-24;  
Matches 44; Conservative 22; Mismatches 41; Indels 5; Gaps 5;  
Db 5 KWCFTLYSSAAREDFLALKEELANYAVGVDEVAPSSGQKHQYLSLKKSIKGLK 64  
QY 16 RWFTLNPNSEDERKKI-REL-PISLFDYFVIGEGNEEGRTPHLOGFANFVKQTFNKV 74  
Db 65 KKY-SSRAHWEARGSDENAKYCSKE-TLILELGFASQSNRKLSEMS 114  
QY 75 W-YLGARCHIEKAKGTDOONKEYCSKEGNLLIEGAPRSQG-QRSDLS 124  
RESULT 11  
ID Q83030 PRELIMINARY; PRT; 285 AA.  
AC Q83030;  
DT 01-NOV-1996 (TREMBLREL.. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL.. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1996 (TREMBLREL.. 01, LAST ANNOTATION UPDATE)  
DE (BBTV.DNA II)  
GN V2.  
OS BANANA BUNCHY TOP VIRUS.  
OC VIRUSES; SSDNA VIRUSES; CIRCOVIRIDAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX WU R.Y., YOU L.R., SOONG T.S.;  
RT SUBMITTED (AUG-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
RL EMBL; L32167; G520792;  
SQ SEQUENCE 285 AA; 32772 MW; F66FA984 CRC32;  
Query Match 8.6%; Score 203; DB 14; Length 285;  
Best Local Similarity 38.8%; Pred. No. 1.65e-20;  
Matches 40; Conservative 18; Mismatches 41; Indels 4; Gaps 4;  
Db 5 KWCFTLYSSAAREDFLALKEEDVHYVGVDEVAPATGQKHQYLSLKKSIKGLK 64  
QY 16 RWFTLNPNSEDERKKI-REL-PISL-FDYFVIGEGNEEGRTPHLOGFANFVKQTFNKV 74  
Db 65 KKY-GSRAHWEIARGSDENAKYCSKE-TLILELGFASQSNRKLSEMS 105  
QY 75 W-YLGARCHIEKAKGTDOONKEYCSKEGNLLIEGAPRSQG 116  
RESULT 12  
ID Q83026 PRELIMINARY; PRT; 286 AA.  
AC Q83026;  
DT 01-NOV-1996 (TREMBLREL.. 01, CREATED)



```

DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE (SMTV DNA I) VI AND C1-C3 GENES, COMPLETE CDS'S.
GN VI.
OS BANANA BUNCHY TOP VIRUS.
OC VIRUSES; SSNA VIRUSES; CIRCOVIRIDAE.
RN [1]
RA SEQUENCE FROM N.A.
RP WU R.Y., YOU L.R., SOONG T.S.;
RL SUBMITTED (AUG-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; L32166; G520787;
SQ SEQUENCE 286 AA; 33249 MW; 52927C5B CRC32;

Query Match      8.3%; Score 136; DB 14; Length 286;
Best Local Similarity 36.5%; Pred. No. 4.58e-19;
Matches 42; Conservative 21; Mismatches 47; Indels 5; Gaps 5;

Db 4 PSLKWCFTLNYSSAERENFLSLKEEDVHVAVGVDEVPATGQKHGQYLSLKKRIRLG 63
QY 13 PKRWVFTLNNPSEDERKKIRELPISLFDYFVIGEGNEEGRTPHLOGFANFVKQTFN 71
Db 64 GLKKKY-GSRAHWEIARTGDENSKYCSKE-TLILELGFVVVNGSNKKRISSEMA 116
QY 72 KVKW-YLGARCHIEKAKTDOONKEYCSKEGNLLIECCAPRSQG-QRSDLTAVS 124

RESULT 13
ID Q86567 PRELIMINARY; PRT; 286 AA.
AC Q86567;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ORF V1 [COMPONENT 1].
OS BANANA BUNCHY TOP VIRUS.
OC VIRUSES; SSNA VIRUSES; CIRCOVIRIDAE.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93187592.
RA HARDING R.M., BURNS T.M., HARNER G., DIETZGEN R.G., DALE J.L.;
RT "Nucleotide sequence of one component of the banana bunchy top virus genome contains a putative replicase gene."
RT J. GEN. VIROL. 74:323-328(1993).
DR EMBL; S56276; G298334;
SQ SEQUENCE 286 AA; 33603 MW; C2FBE5C7 CRC32;

Query Match      5.8%; Score 136; DB 14; Length 286;
Best Local Similarity 31.0%; Pred. No. 1.96e-07;
Matches 26; Conservative 27; Mismatches 23; Indels 8; Gaps 8;

Db 8 WMFTINNP-TLPV-MRD-EIK-YMYQV-ERG-QEG-TRHVQGYVEMKRSSLKQMRGF 60
QY 17 WVFTLNPSEDERKKIRELPISLFDYFVIGEGNEEGRTPHLOGFANFVKQTFN 76
Db 61 PGA-HLEKRGSOEARSYCMKE 83
QY 77 LGARCHIEKAKTDOONKEYCSKE 100

RESULT 14
ID Q65378 PRELIMINARY; PRT; 286 AA.
AC Q65378;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 33.5 KD PROTEIN.
OS BANANA BUNCHY TOP VIRUS.
OC VIRUSES; SSNA VIRUSES; CIRCOVIRIDAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HAWAIIAN;
RA XIE W.S., HU J.S.;
RL PHYTOPATHOLOGY 0:0-0(0).
DR EMBL; U18077; G603196;

```

```

KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 286 AA; 33488 MW; 3C722F3E CRC32;

Query Match      5.8%; Score 136; DB 14; Length 286;
Best Local Similarity 31.0%; Pred. No. 1.96e-07;
Matches 26; Conservative 27; Mismatches 23; Indels 8; Gaps 8;

Db 8 WMFTINNP-TLPV-MRD-EIK-YMYQV-ERG-QEG-TRHVQGYVEMKRSSLKQMRGF 60
QY 17 WVFTLNPSEDERKKIRELPISLFDYFVIGEGNEEGRTPHLOGFANFVKQTFN 76
Db 61 PGA-HLEKRGSOEARSYCMKE 83
QY 77 LGARCHIEKAKTDOONKEYCSKE 100

RESULT 15
ID Q70710 PRELIMINARY; PRT; 3085 AA.
AC Q70710;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE POLYPROTEIN.
OS INFECTIOUS FLACHERIE VIRUS.
OC VIRUSES; SSNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98166871.
RA ISAWA H., ASANO S., SAHARA K., IIZUKA T., BANDO H.;
RT "Analysis of genetic information of an insect picorna-like virus, infectious flacherie virus of silkworm: evidence for evolutionary relationships among insect, mammalian and plant picorna(-like) virus-s."
RT ARCH. VIROL. 143:127-143(1998).
DR EMBL; AB000905; D1026299;
KW POLYPROTEIN.
SQ SEQUENCE 3085 AA; 345798 MW; EEB514A1 CRC32;

Query Match      5.5%; Score 130; DB 14; Length 3085;
Best Local Similarity 35.3%; Pred. No. 2.28e-06;
Matches 12; Conservative 13; Mismatches 9; Indels 0; Gaps 0;

Db 1411 PGKFEHPNPVTRNPTSQWNGINGOPIVLYDDI 1444
QY 184 AAFANPETTYKPKKXWDGKHGEKVVVIDDF 217

Search completed: Wed Dec 22 10:13:43 1999
Job time : 28 secs.

```

\*\*\*\*\*  
WATERMAN  
\*\*\*\*\*  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Wed Dec 22 10:29:55 1999; MasPar time 9.99 Seconds  
Tabular output not generated. -289.470 Million cell updates/sec

Title: >US-09-209-961-21  
Description: (1-53) from US09209961.pep  
Perfect Score: 416  
Sequence: 1 MVFIIHLGFKWGFKIKFSE.....VFTVFERSAEYVYVHSRGL 53

Scoring table: PAM 150  
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: sptrembl9  
1:sp.archaea 2:sp.bacteria 3:sp.fungi 4:sp.human  
5:sp.invertebrate 6:sp.mammal 7:sp.mhc 8:sp.organelle  
9:sp.phase 10:sp.plant 11:sp.rodent 12:sp.unclassified  
13:sp.vertebrate 14:sp.virus

Statistics: Mean 37.523; Variance 83.063; scale 0.452

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	416	100.0	53	14	093201 P6.2.	4.88e-54
2	412	99.0	53	14	091861 ORF5.	2.55e-53
3	406	97.6	53	14	056128 STRAIN PMWS PCV, COMPL	3.04e-52
4	95	22.8	124	2	086225 HYPOTHETICAL 14.5 KD P	3.88e-01
5	92	22.1	1199	14	083066 DNA-DEPENDENT RNA POLY	9.18e-01
6	91	21.9	292	2	085245 ABC TRANSPORTER.	1.22e+00
7	90	21.6	236	8	063672 NADH DEHYDROGENASE SUB	1.62e+00
8	90	21.6	236	8	063673 NADH DEHYDROGENASE SUB	1.62e+00
9	89	21.4	302	2	050501 PROBABLE TRANSMEMBRANE	1.14e+00
10	87	20.9	276	2	P72880 ABC TRANSPORTER.	3.72e+00
11	87	20.9	1268	5	Q20335 PGP-3 PROTEIN.	3.72e+00
12	86	20.7	337	5	045805 T23D5.6 PROTEIN.	4.90e+00
13	86	20.7	434	10	082589 F1104.6 PROTEIN.	4.90e+00
14	85	20.4	315	1	029337 HYPOTHETICAL 36.7 KD P	6.43e+00
15	85	20.4	377	5	021936 R11D1.5 PROTEIN.	6.43e+00
16	84	20.2	152	8	096097 NADH DEHYDROGENASE SUB	8.43e+00
17	84	20.2	325	8	078874 NADH DEHYDROGENASE SUB	8.43e+00
18	84	20.2	422	2	P95734 FEM.	8.43e+00
19	84	20.2	438	10	Q40017 ALPHA-AMYLASE 2.	8.43e+00
20	84	20.2	477	3	Q06134 CHROMOSOME XII COSMID	8.43e+00

21	84	20.2	924	2	006842 B14, PUTATIVE POLYGALA	8.43e+00
22	83	20.0	106	1	006101 HYPOTHETICAL 12.8 KD P	1.10e+01
23	83	20.0	120	2	068285 LARGE CONDUCTANCE MECH	1.10e+01
24	83	20.0	236	8	063674 NADH DEHYDROGENASE SUB	1.10e+01
25	83	20.0	236	8	063671 NADH DEHYDROGENASE SUB	1.10e+01
26	83	20.0	236	8	063944 NADH DEHYDROGENASE SUB	1.10e+01
27	83	20.0	236	8	063670 NADH DEHYDROGENASE SUB	1.10e+01
28	83	20.0	236	8	063666 NADH DEHYDROGENASE SUB	1.10e+01
29	83	20.0	299	5	044810 F14D2.1 PROTEIN.	1.10e+01
30	82	19.7	100	3	087321 ORF YJ025C-A.	1.44e+01
31	82	19.7	152	8	096091 NADH DEHYDROGENASE SUB	1.44e+01
32	82	19.7	152	8	095830 NADH DEHYDROGENASE SUB	1.44e+01
33	82	19.7	209	2	095773 CADB.	1.44e+01
34	82	19.7	236	8	063969 NADH DEHYDROGENASE SUB	1.44e+01
35	82	19.7	236	8	063668 NADH DEHYDROGENASE SUB	1.44e+01
36	82	19.7	299	5	P91070 SIMILARITY TO C. ELEGA	1.44e+01
37	82	19.7	383	1	026855 CONSERVED PROTEIN.	1.44e+01
38	82	19.7	489	5	093704 F42D1.3 PROTEIN.	1.44e+01
39	82	19.7	528	10	080838 PUTATIVE ATP-DEPENDENT	1.44e+01
40	82	19.7	840	3	005848 J1575L ORF.	1.44e+01
41	82	19.7	1435	11	088922 CL2BB.	1.44e+01
42	82	19.7	1452	11	088918 CL2AA.	1.44e+01
43	82	19.7	1457	11	088921 CL2BA.	1.44e+01
44	82	19.7	1478	11	088923 CL2BC.	1.44e+01
45	82	19.7	2035	4	092508 MYELOBLAST KIAA0233.	1.44e+01

ALIGNMENTS

RESULT	ID	PRELIMINARY	PRT	53 AA
AC	093201			
DE	01-NOV-1999 (TREMREL. 08, CREATED)			
DE	01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)			
DE	01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)			
DE	P6.2.			
OS	PORCINE CIRCOVIRUS.			
OC	VIRUSES; 3SDNA VIRUSES; CIRCOVIRIDAE; CIRCOVIRUS.			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RP	STRAIN-412;			
RA	WANG L., WILLSON P., CHOW B., GIBBONS E., BABIUK L.;			
RA	"Emergence of a new porcine circovirus.;"			
RL	SUBMITTED (AUG-1998) TO EMBL/GENBANK/DBJ DATA BANKS.			
YN	(2)			
RP	SEQUENCE FROM N.A.			
FC	STRAIN-PORCINE CIRCOVIRUS TYPE II;			
RX	MEDLINE: 98418498.			
RA	MEEHAN B.M., MCNEILLY F.M., TODD D., KENNEDY S., JEWHRST V.,			
RA	ELLIS J.A., HASSARD L.E., CLARK E.G., HAINES D.M., ALLAN G.M.;			
RT	"Characterisation of novel circovirus DNAs associated with wasting			
RT	syndromes in pigs.;"			
RL	J. GEN VIROL. 79:2171-2199(1998).			
RN	(3)			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-PORCINE CIRCOVIRUS TYPE II;			
RA	MEEHAN B.M.;			
PI	SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.			
NR	EMBL: AF085695; G3668366;			
NR	EMBL: AF055392; G3598813;			
SO	SEQUENCE 53 AA; 6176 MW; 59859B1D CRC32;			
Query Match 100.08; Score 416; DB 14; Length 53;				
Besc Local Similarity 100.08; Pred. No. 4.88e-54;				
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
DB	1	MVFIIHLGFKWGFKIKFSELYTHGTDIVLVVFTVFERSAEYVYVHSRGL 53		
QY	1	MVFIIHLGFKWGFKIKFSELYTHGTDIVLVVFTVFERSAEYVYVHSRGL 53		
RESULT	2			
ID	091861	PRELIMINARY	PRT	53 AA.

```

AC O91861;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ORF5.
OS PORCINE CIRCOVIRUS.
OC VIRUSES; SSDNA VIRUSES; CIRCOVIRIDAE; CIRCOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PORCINE CIRCOVIRUS TYPE II;
RX MEDLINE; 98418498.
RA MEEHAN B.M., MCNEILLY F.M., TODD D., KENNEDY S., JEWURST V.,
RA ELLIS J.A., HASSARD L.E., CLARK E.G., HAINES D.M., ALLAN G.M.;
RT "Characterisation of novel circovirus DNAs associated with wasting
RT syndromes in pigs.";
RL J. GEN. VIROL. 79:2171-2199(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-PORCINE CIRCOVIRUS TYPE II;
RA MEEHAN B.M.;
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF055391; G3598801;
SQ SEQUENCE 53 AA; 6192 MW; 095836D9 CRC32;

Query Match
Best Local Similarity 99.08; Score 412; DB 14; Length 53;
Matches 52; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 1 MFIIHLGFKWGVKFKFSELYIHGTYDVLVVFVFFERSAEAYVHISRGL 53
QY 1 MFIIHLGFKWGVKFKFSELYIHGTYDVLVVFVFFERSAEAYVHISRGL 53

RESULT 3
ID O56128; PRELIMINARY; PRT; 53 AA.
AC O56128;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE STRAIN PMWS PCV, COMPLETE GENOME.
OS PORCINE CIRCOVIRUS.
OC VIRUSES; SSDNA VIRUSES; CIRCOVIRIDAE; CIRCOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PMWS PCV;
RX MEDLINE; 98241772.
RA HAMEL A.L., LIN L.L., NAVAR G.P.S.;
RT "Nucleotide sequence of porcine circovirus associated with
RT postweaning multisystemic wasting syndrome in pigs.";
RL J. VIROL. 72:5262-5267(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-PMWS PCV;
RA HAMEL A.L., LIN L.L., NAVAR G.P.S.;
RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF027217; G2689650;
SQ SEQUENCE 53 AA; 6226 MW; B2D59A3C CRC32;

Query Match
Best Local Similarity 97.68; Score 406; DB 14; Length 53;
Matches 51; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 1 MFIIHLGFKWGVKFKFSELYIHGTYDVLVVFVFFERSAEAYVHISRGL 53
QY 1 MFIIHLGFKWGVKFKFSELYIHGTYDVLVVFVFFERSAEAYVHISRGL 53

RESULT 4
ID O86225; PRELIMINARY; PRT; 124 AA.
AC O86225;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)

```

```

DE HYPOTHETICAL 14.5 KD PROTEIN.
GN H10485.1.
OS HAEMOPHILUS INFLUENZAE RD.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURACEAE;
OC HAEMOPHILUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95350630.
RA FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,
RA KERLAVAGE A.R., BULT C.J., TOMB J., DOUGHERTY B.A., MERRICK J.M.,
RA MCKENNEY K., SUTTON G.G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,
RA SCOTT J.D., SHIRLEY R., LIU L.I., GLODEK A., KELLEY J.M.,
RA WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLOM E., COTTON M.D.,
RA UTERBACK T., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,
RA FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEOGHAGEN N.S., GNEHM C.L.,
RA McDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O., VENTER J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL SCIENCE 269:496-512(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96398784.
RA TATSOV R.L.; MUSHEGIAN A.R., BORK P., BROWN N.P., HAYES W.S.,
RA BORODOVSKY M., RUDD K.E., KOONIN E.V.;
RT "Metabolism and evolution of Haemophilus influenzae deduced from a
RT whole-genome comparison with Escherichia coli.";
RL CURR. BIOL. 6:279-291(1996).
RN [3]
RP SEQUENCE FROM N.A.
RA WHITE O., CLAYTON R.A., KERLAVAGE A.R., FLEISCHMANN R.D.;
RL SUBMITTED (JUL-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [4]
RP SEQUENCE FROM N.A.
RA WHITE O., CLAYTON R.A., KERLAVAGE A.R., FLEISCHMANN R.D.;
RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [5]
RP SEQUENCE FROM N.A.
RA WHITE O., CLAYTON R.A., KERLAVAGE A.R., FLEISCHMANN R.D., PETERSON J.,
RA HICKEY E., DODSON R., GWINN M.;
RL SUBMITTED (MAY-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U32730; G3212192;
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 124 AA; 14546 MW; 014F579C CRC32;

Query Match
Best Local Similarity 22.8%; Score 95; DB 2; Length 124;
Matches 13; Conservative 10; Mismatches 6; Indels 3; Gaps 3;

Db 85 MVFII-TAIKW-LFVINFIATFV-GFLALVYL 113
QY 1 MVFIIHLGFKWGVKFKFSELYIHGTYDVLV 32

RESULT 5
ID Q83066; PRELIMINARY; PRT; 1199 AA.
AC Q83066;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE DNA-DEPENDENT RNA POLYMERASE LARGEST SUBUNIT.
OS LYMPHOCYSTIS DISEASE VIRUS OF FISH.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; IRIDOVIRIDAE; LYMPHOCYSTIVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RA MOELLER M., SCHNITZLER P., KOONIN E.V., DARAI G.;
RL J. GEN. VIROL. 0:0-0(0).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95248278.
RA MULLER M., SCHNITZLER P., KOONIN E.V., DARAI G.;
RT "Identification and properties of the largest subunit of the
RT DNA-dependent RNA polymerase of fish lymphocystis disease virus:
RT dramatic difference in the domain organization in the family

```

```
RT Iridoviridae.
RL J. GEN. VIROL. 76:1099-1107(1995).
DR EMBL; L34213; G755032;
DR PFAM; PF00623; RNA_POL_A; 1.
SQ SEQUENCE 1199 AA; 134368 MW; 74BBF357 CRC32;

Query Match 22.1%; Score 92; DB 14; Length 1199;
Best Local Similarity 28.0%; Pred. No. 9.18e-01;
Matches 14; Conservative 15; Mismatches 18; Indels 3; Gaps 2;

Db 763 IPLCFKLAYKLNLYTGLPTNPNNLELFFKQLPPIQYVQLQHR 832
Y 5 IHLGFKWGVFKIK--FSE-LYIHGTYDIVLVVFTVFERSAEAYVHISR 51

RESULT 6
ID O85245 PRELIMINARY; PRT; 292 AA.
AC O85245
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ABC TRANSPORTER.
OS THERMOTOGA NEAPOLITANA.
OC BACTERIA; THERMOTOGALES; THERMOTOGA.
RN [1]
RP SEQUENCE FROM N.A.
RA SWATEK G.C., YERNOOL D.A., KING M., CHASSY B., EVELEIGH D.E.;
RT "Molecular Characterization of a galactose utilization operon in
RT Thermotoga neapolitana."
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF055482; G3252893;
SQ SEQUENCE 292 AA; 33362 MW; 1D51E4ED CRC32;

Query Match 21.9%; Score 91; DB 2; Length 292;
Best Local Similarity 27.7%; Pred. No. 1.22e+00;
Matches 13; Conservative 18; Mismatches 12; Indels 4; Gaps 4;

Db 5 WGVFKDLFRDYFTVAFIVLVLLIFLSVLSHSPDPIRIYVPRGL 51
Y 11 WGVFKFSEL-YINGTYDIVLVVFTVFER-SA-EAYVW-HISRGL 53

RESULT 7
ID O63672 PRELIMINARY; PRT; 236 AA.
AC O63672
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE NADH DEHYDROGENASE SUBUNIT 4 (FRAGMENT).
GN NDA.
OS CALLITHRIX JACCHUS (COMMON MARMOSET).
OC MITOCHONDRION.
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC PLATYRRHINI; CALLITRICHIDAE; CALLITHRIX.
RN [1]
RP SEQUENCE FROM N.A.
RA PASTORINI J., FORSTNER M.R.J., MARTIN R.D., MELNICK D.J.;
RA J. MOL. EVOL. 0:0-0(1998).
DR EMBL; AF053695; G2995984;
FT NON TER
SQ SEQUENCE 236 AA; 26193 MW; F7249638 CRC32;

Query Match 21.6%; Score 90; DB 8; Length 236;
Best Local Similarity 26.9%; Pred. No. 1.62e+00;
Matches 14; Conservative 14; Mismatches 23; Indels 1; Gaps 1;

Db 156 LITILASFSNFTIMFTGLNML-ITALYSLHMTSTQRPITYSTSNVKPL 206
Y 2 VFIIHLGFKWGVFKIKFSELYIHGTYDIVLVVFTVFERSAEAYVHISRGL 53

RT Iridoviridae.
RL J. GEN. VIROL. 76:1099-1107(1995).
DR EMBL; L34213; G755032;
DR PFAM; PF00623; RNA_POL_A; 1.
SQ SEQUENCE 1199 AA; 134368 MW; 74BBF357 CRC32;

Query Match 22.1%; Score 92; DB 14; Length 1199;
Best Local Similarity 28.0%; Pred. No. 9.18e-01;
Matches 14; Conservative 15; Mismatches 18; Indels 3; Gaps 2;

Db 763 IPLCFKLAYKLNLYTGLPTNPNNLELFFKQLPPIQYVQLQHR 832
Y 5 IHLGFKWGVFKIK--FSE-LYIHGTYDIVLVVFTVFERSAEAYVHISR 51

RESULT 6
ID O85245 PRELIMINARY; PRT; 292 AA.
AC O85245
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ABC TRANSPORTER.
OS THERMOTOGA NEAPOLITANA.
OC BACTERIA; THERMOTOGALES; THERMOTOGA.
RN [1]
RP SEQUENCE FROM N.A.
RA SWATEK G.C., YERNOOL D.A., KING M., CHASSY B., EVELEIGH D.E.;
RT "Molecular Characterization of a galactose utilization operon in
RT Thermotoga neapolitana."
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF055482; G3252893;
SQ SEQUENCE 292 AA; 33362 MW; 1D51E4ED CRC32;

Query Match 21.9%; Score 91; DB 2; Length 292;
Best Local Similarity 27.7%; Pred. No. 1.22e+00;
Matches 13; Conservative 18; Mismatches 12; Indels 4; Gaps 4;

Db 5 WGVFKDLFRDYFTVAFIVLVLLIFLSVLSHSPDPIRIYVPRGL 51
Y 11 WGVFKFSEL-YINGTYDIVLVVFTVFER-SA-EAYVW-HISRGL 53

RESULT 7
ID O63672 PRELIMINARY; PRT; 236 AA.
AC O63672
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE NADH DEHYDROGENASE SUBUNIT 4 (FRAGMENT).
GN NDA.
OS CALLITHRIX JACCHUS (COMMON MARMOSET).
OC MITOCHONDRION.
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC PLATYRRHINI; CALLITRICHIDAE; CALLITHRIX.
RN [1]
RP SEQUENCE FROM N.A.
RA PASTORINI J., FORSTNER M.R.J., MARTIN R.D., MELNICK D.J.;
RA J. MOL. EVOL. 0:0-0(1998).
DR EMBL; AF053695; G2995984;
FT NON TER
SQ SEQUENCE 236 AA; 26193 MW; F7249638 CRC32;

Query Match 21.6%; Score 90; DB 8; Length 236;
Best Local Similarity 26.9%; Pred. No. 1.62e+00;
Matches 14; Conservative 14; Mismatches 23; Indels 1; Gaps 1;

Db 156 LITILASFSNFTIMFTGLNML-ITALYSLHMTSTQRPITYSTSNVKPL 206
Y 2 VFIIHLGFKWGVFKIKFSELYIHGTYDIVLVVFTVFERSAEAYVHISRGL 53
```

```
RESULT 8
ID O63673 PRELIMINARY; PRT; 236 AA.
AC O63673
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE NADH DEHYDROGENASE SUBUNIT 4 (FRAGMENT).
GN NDA.
OS CALLITHRIX JACCHUS (COMMON MARMOSET).
OC MITOCHONDRION.
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC PLATYRRHINI; CALLITRICHIDAE; CALLITHRIX.
RN [1]
RP SEQUENCE FROM N.A.
RA PASTORINI J., FORSTNER M.R.J., MARTIN R.D., MELNICK D.J.;
RA J. MOL. EVOL. 0:0-0(1998).
DR EMBL; AF053696; G2995986;
FT NON TER
SQ SEQUENCE 236 AA; 26209 MW; 7C74B180 CRC32;

Query Match 21.6%; Score 90; DB 8; Length 236;
Best Local Similarity 26.9%; Pred. No. 1.62e+00;
Matches 14; Conservative 14; Mismatches 23; Indels 1; Gaps 1;

Db 156 LITILASFSNFTIMFTGLNML-ITALYSLHMTSTQRPITYSTSNVKPL 206
Y 2 VFIIHLGFKWGVFKIKFSELYIHGTYDIVLVVFTVFERSAEAYVHISRGL 53

RESULT 9
ID O50501 PRELIMINARY; PRT; 302 AA.
AC O50501
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PROBABLE TRANSMEMBRANE TRANSPORT PROTEIN.
GN SC7B7.04.
OS STREPTOMYCES COELICOLOR.
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC ACTINOMYCETALES; STREPTOMYCINAE; STREPTOMYCETACEAE; STREPTOMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RA MORPHY L., HARRIS D.;
RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN-A3(2);
RA PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RA STRAIN-A3(2);
RA MEDLINE; 97000351.
RA REDENBACH M., KIESER H.M., DENAPATE D., EICHNER A., CULLUM J.,
RA KINASHI H., HOPWOOD D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb streptomycetes coelicolor A3(2) chromosome."
RL MOL. MICROBIOL. 21:77-96(1996).
JR EMBL; AL000199; EI202318;
DR PROSITE; PS00402; BPD_TRANSF_INN_MEMBR; 1.
RW TRANSMEMBRANE; TRANSPORT.
SQ SEQUENCE 302 AA; 33763 MW; 73A28012 CRC32;

Query Match 21.4%; Score 89; DB 2; Length 302;
Best Local Similarity 29.7%; Pred. No. 2.14e+00;
Matches 11; Conservative 10; Mismatches 15; Indels 1; Gaps 1;

Db 262 VYDIYGFALFAAIVVTPVLLVYCVFQRIAGSVQ 298
Y 13 VFAJFSELYIHGTYDIVLVVFTVFERSAEAYVH 48
```





\*\*\*\*\*  
M P S R C H  
\*\*\*\*\*  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Blocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Dec 22 10:17:08 1999; MasPar time 19.80 Seconds  
642.264 Million cell updates/sec

Tabular output not generated.

Title: >US-09-209-961-5  
Description: (1-233) from US09209961.pep  
Perfect Score: 1805  
Sequence: 1 MTPRRRRRRRRRSHLG.....RVTMYQVREFNKLDPPLP 233

Scoring table: PAM 150  
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: sptrembl9  
1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phase 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 45.232; Variance 85.435; scale 0.529

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Description	Pred. No.
1	1795	99.4	233 14 091865 ORF2.	0.00e+00
2	1746	96.7	233 14 091862 ORF2.	0.00e+00
3	1746	96.7	233 14 056129 STRAIN PMWS PCV, COMPL	0.00e+00
4	1729	95.8	233 14 091868 ORF2.	0.00e+00
5	1196	66.3	230 14 090239 CAPSID PROTEIN.	1.65e-228
6	233	12.9	244 14 090237 CAPSID PROTEIN (FRAGME	2.59e-23
7	146	8.1	770 14 070796 LONG ORF.	6.62e-08
8	143	7.9	770 14 070796 LONG ORF.	2.03e-07
9	143	7.9	770 14 070739 ORF1.	2.03e-07
10	140	7.8	768 14 070810 LONG ORF.	6.17e-07
11	139	7.7	770 14 090364 ORF1 PROTEIN.	8.91e-07
12	133	7.4	769 14 070804 LONG ORF.	7.91e-06
13	133	7.4	769 14 070802 LONG ORF.	7.91e-06
14	133	7.4	769 14 070798 LONG ORF.	7.91e-06
15	133	7.4	770 14 070794 LONG ORF.	7.91e-06
16	127	7.0	124 14 070806 LONG ORF (FRAGMENT).	6.72e-05
17	125	6.9	769 14 070808 LONG ORF.	1.36e-04
18	120	6.6	134 2 P96909 HYPOTHETICAL 15.1 KD P	7.66e-04
19	120	6.6	1538 2 O53395 PGRS-FAMILY PROTEIN.	7.66e-04
20	115	6.4	823 11 O88378 PRP4 PROTEIN KINASE HO	4.16e-03

21	113	6.3	62	6	018749	PROTAMINE P1.	8.11e-03
22	114	6.3	113	3	Q08428	ORF YOR053W.	5.82e-03
23	110	6.1	126	3	Q12444	ORF YOR309C.	2.17e-02
24	110	6.1	586	11	P97358	TAFI68.	2.17e-02
25	109	6.0	225	1	O58203	225AA LONG HYPOTHETICA	3.01e-02
26	106	5.9	171	14	O71097	PVII	7.90e-02
27	104	5.8	200	14	Q98691	COAT PROTEIN.	1.49e-01
28	104	5.8	327	5	O02394	DROSOPHILIN MELANOGAST	1.49e-01
29	103	5.7	344	5	O16526	T05H4.1 PROTEIN.	2.04e-01
30	102	5.7	449	14	Q96573	NUCLEOCAPSID PROTEIN V	2.79e-01
31	102	5.7	1288	4	O15038	KIAA0324 (FRAGMENT).	2.79e-01
32	102	5.7	1791	4	O60382	KIAA0324 (FRAGMENT).	2.79e-01
33	101	5.6	271	10	Q26300	TRANSCRIPTION FACTOR.	3.80e-01
34	101	5.6	710	3	Q00210	CHROMOSOME SEGREGATION	3.80e-01
35	101	5.6	968	4	O60289	KIAA0542 PROTEIN.	3.80e-01
36	101	5.6	1473	11	Q63625	CTD-BINDING SR-LIKE PR	3.80e-01
37	100	5.5	111	10	Q39682	GLYCINE-RICH PROTEIN (	5.17e-01
38	99	5.5	153	10	Q39680	HEAT-SHOCK COGNATE (FR	7.02e-01
39	100	5.5	163	4	Q15410	NUCLEIC ACID BINDING P	5.17e-01
40	95	5.5	302	3	Q05330	POTATIVE ORF.	7.02e-01
41	100	5.5	449	14	Q96572	COMPLETE GENOME: VP2,	5.17e-01
42	100	5.5	449	14	O41449	52 KDA MAJOR CAPSID PR	5.17e-01
43	99	5.5	474	2	P75295	P02_ORF474 PROTEIN.	7.02e-01
44	98	5.4	249	2	O83299	UDP-N-ACETYL-D-MANNOSA	9.51e-01
45	98	5.4	2183	14	Q86887	POLYPROTEIN.	9.51e-01

ALIGNMENTS

RESULT	1	PRELIMINARY;	PRT;	233 AA.
ID	O91865			
AC	O91865;			
DT	01-NOV-1998 (TREMREL. 08, CREATED)			
DT	01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)			
DT	01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)			
DE	ORF2.			
OS	PORCINE CIRCOVIRUS.			
OC	VIRUSES; SSDNA VIRUSES; CIRCOVIRIDAE; CIRCOVIRUS.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-PORCINE CIRCOVIRUS TYPE II;			
RX	MEDLINE; 98418498			
RA	MECHAN B.M., MCNEILLY F.M., TODD D., KENNEDY S., JEWTHURST V.,			
PA	ELIJS J.A., HASSARD L.E., CLARK E.G., HAINES D.M., ALLAN G.M.;			
FT	"Characterisation of novel circovirus DNAs associated with wasting			
XI	syndromes in pigs."			
WJ	J. GEN. VIROL. 79:2171-2159(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-PORCINE CIRCOVIRUS TYPE II;			
RA	MECHAN B.M.;			
RL	SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.			
DR	EMBL; AF355392; G3598810;			
EQ	SEQUENCE 233 AA; 27840 MW; 721197FF CRC32;			
Query Match 99.4%; Score 1795; DB 14; Length 233;				
Best Local Similarity 99.1%; Pred. No. 0.00e+00;				
Matches: 233; Conservative 1; Mismatches 1; Indels 0; Gaps 0;				
LQ	1	MTYPRRRRRRRRSHLGQILRRPWLPHRPHRYRWRKNGIFNTRLSRTFGYTVKRT	60	
QY	1	MTYPRRRRRRRRSHLGQILRRPWLPHRPHRYRWRKNGIFNTRLSRTFGYTVKRT	60	
DF	61	TVTTPSWAVDMRFRKIDDPVPPGGGPNKISIPFEYIRIRKRVKVEFWPCSPITGDRGVGS	120	
QY	61	TVTTPSWAVDMRFRKIDDPVPPGGGPNKISIPFEYIRIRKRVKVEFWPCSPITGDRGVGS	120	
FD	121	TAVILDDNFTKATLTYDPYNYSSRRHPIQPFYSHSRYFTPKPVLDSIDTFQPNKR	180	
QY	121	TAVILDDNFTKATLTYDPYNYSSRRHPIQPFYSHSRYFTPKPVLDSIDTFQPNKR	180	
b	181	NOLWLRLOTSGNVDHVGAAAFENSKYDODYINRVTMYQVREFNKLDPPLP	233	

QY 181 NQLWRLQTSNVDHVGGLTAFENSKYDQDYNIRVTMYVQFREFNLKDPPLP 233

## RESULT 2

ID O91862; PRELIMINARY; PRT: 233 AA.  
 AC O91862;  
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE ORF2.  
 OS PORCINE CIRCOVIRUS.  
 OC VIRUSES; SSDNA VIRUSES; CIRCOVIRIDAE; CIRCOVIRUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-PORCINE CIRCOVIRUS TYPE II;  
 RX MEDLINE; 98418498.  
 RA MEEHAN B.M., MCNEILLY F.M., TODD D., KENNEDY S., JEWURST V.,  
 RA ELLIS J.A., HASSARD L.E., CLARK E.G., HAINES D.M., ALLAN G.M.;  
 RT "Characterisation of novel circovirus DNAs associated with wasting  
 RT syndromes in pigs.";  
 RL J. GEN. VIROL. 79:2171-2199(1998).  
 RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-PORCINE CIRCOVIRUS TYPE II;

RA MEEHAN B.M.;

RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL; AF055391; G3598798; -

SQ SEQUENCE 233 AA; 27882 MW; 8C4A18BA CRC32;

Query Match 96.7%; Score 1746; DB 14; Length 233;

Best Local Similarity 97.0%; Pred. No. 0.00e+00;

Matches 226; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Db 1 MTPRRYRRRRHPRSHLQGLRRRRLVHPHRYRWRKNGINFRLSRTFGYTVKAT 60  
 QY 1 MTPRRYRRRRHPRSHLQGLRRRRLVHPHRYRWRKNGINFRLSRTFGYTVKRT 60  
 Db 61 TVRTPSWAVDMRNIDDFPPGGGTNKSIPPEYIRKVKVEFWPCSPITQDGRVGS 120  
 QY 61 TVTTPSWAVDMRMFKIDDFFPPGGGTNKSIPPEYIRKVKVEFWPCSPITQDGRVGS 120  
 Db 121 TAVILDDNFVTKATLYDYPVYSSRHTIPQPSYHSRYFTPKPVLDSIDYFQPNKR 180  
 QY 121 TAVILDDNFVTKATLYDYPVYSSRHTIPQPSYHSRYFTPKPVLDSIDYFQPNKR 180  
 Db 181 NQLWRLQTSNVDHVGGLTAFENSKYDQDYNIRVTMYVQFREFNLKDPPLK 233  
 QY 181 NQLWRLQTSNVDHVGGLTAFENSKYDQDYNIRVTMYVQFREFNLKDPPLP 233

## RESULT 3

ID O56129; PRELIMINARY; PRT: 233 AA.  
 AC O56129;  
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE STRAIN PMWS PCV, COMPLETE GENOME.  
 OS PORCINE CIRCOVIRUS.  
 OC VIRUSES; SSDNA VIRUSES; CIRCOVIRIDAE; CIRCOVIRUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-PMWS PCV;  
 RX MEDLINE; 98241772.  
 RA HAMEL A.L., LIN L.L., NAVAR G.P.S.;  
 RA "Nucleotide sequence of porcine circovirus associated with  
 RT postweaning multisystemic wasting syndrome in pigs.";  
 RL J. VIROL. 72:5262-5267(1998).  
 RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-PMWS PCV;

RA HAMEL A.L., LIN L.L., NAVAR G.P.S.;

RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL; AF027217; G2689647; -

SQ SEQUENCE 233 AA; 27897 MW; CB1E261D CRC32;

Query Match 96.7%; Score 1746; DB 14; Length 233;

Best Local Similarity 97.0%; Pred. No. 0.00e+00;

Matches 226; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Db 1 MTPRRYRRRRHPRSHLQGLRRRRLVHPHRYRWRKNGINFRLSRTFGYTVKAT 60  
 QY 1 MTPRRYRRRRHPRSHLQGLRRRRLVHPHRYRWRKNGINFRLSRTFGYTVKRT 60  
 Db 61 TVRTPSWAVDMRNIDDFPPGGGTNKSIPPEYIRKVKVEFWPCSPITQDGRVGS 120  
 QY 61 TVTTPSWAVDMRMFKIDDFFPPGGGTNKSIPPEYIRKVKVEFWPCSPITQDGRVGS 120  
 Db 121 TAVILDDNFVTKATLYDYPVYSSRHTIPQPSYHSRYFTPKPVLDSIDYFQPNKR 180  
 QY 121 TAVILDDNFVTKATLYDYPVYSSRHTIPQPSYHSRYFTPKPVLDSIDYFQPNKR 180  
 Db 181 NQLWRLQTSNVDHVGGLTAFENSKYDQDYNIRVTMYVQFREFNLKDPPLK 233  
 QY 181 NQLWRLQTSNVDHVGGLTAFENSKYDQDYNIRVTMYVQFREFNLKDPPLP 233

## RESULT 4

ID O91868; PRELIMINARY; PRT: 233 AA.  
 AC O91868;  
 DT 01-NOV-1998 (TREMBLREL. 08, CREATED)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE ORF2  
 OS PORCINE CIRCOVIRUS.  
 OC VIRUSES; SSDNA VIRUSES; CIRCOVIRIDAE; CIRCOVIRUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-PORCINE CIRCOVIRUS TYPE II;  
 RX MEDLINE; 98418498.  
 RA MEEHAN B.M., MCNEILLY F.M., TODD D., KENNEDY S., JEWURST V.,  
 RA ELLIS J.A., HASSARD L.E., CLARK E.G., HAINES D.M., ALLAN G.M.;  
 RT "Characterisation of novel circovirus DNAs associated with wasting  
 RT syndromes in pigs.";  
 RL J. GEN. VIROL. 79:2171-2199(1998).  
 RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-PORCINE CIRCOVIRUS TYPE II;

RA MEEHAN B.M.;

RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL; AF055393; G3598822; -

SQ SEQUENCE 233 AA; 27867 MW; A6EBC19 CRC32;

Query Match 95.8%; Score 1729; DB 14; Length 233;

Best Local Similarity 94.0%; Pred. No. 0.00e+00;

Matches 219; Conservative 9; Mismatches 5; Indels 0; Gaps 0;

Db 1 MTPRRYRRRRHPRSHLQGLRRRRLVHPHRYRWRKNGINFRLSRTFGYTVKRT 60  
 QY 1 MTPRRYRRRRHPRSHLQGLRRRRLVHPHRYRWRKNGINFRLSRTFGYTVKRT 60  
 Db 61 TVRTPSWAVDMRNIDDFPPGGGTNKSIPPEYIRKVKVEFWPCSPITQDGRVGS 120  
 QY 61 TVTTPSWAVDMRMFKIDDFFPPGGGTNKSIPPEYIRKVKVEFWPCSPITQDGRVGS 120  
 Db 121 SAVILDDNFVTKATLYDYPVYSSRHTIPQPSYHSRYFTPKPVLDSIDYFQPNKR 180  
 QY 121 TAVILDDNFVTKATLYDYPVYSSRHTIPQPSYHSRYFTPKPVLDSIDYFQPNKR 180  
 Db 181 NQLWRLQTSNVDHVGGLTAFENSKYDQDYNIRVTMYVQFREFNLKDPPLN 233  
 QY 181 NQLWRLQTSNVDHVGGLTAFENSKYDQDYNIRVTMYVQFREFNLKDPPLP 233

## RESULT 5

ID O90239; PRELIMINARY; PRT: 230 AA.  
 AC O90239;



```

DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE CAPSID PROTEIN.
GN C1.
OS PORCINE CIRCOVIRUS.
OC VIRUSES; SSDNA VIRUSES; CIRCOVIRIDAE; CIRCOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 99003925.
RA NIAGRO F.D., FORSTHOEFFEL A.N., LAWTHOR R.P., KAMALANATHAN L.,
RITCHIE B.W., LATIMER K.S., LUKERT P.D.;
RT "Beak and feather disease virus and porcine circovirus genomes:
intermediates between the geminiviruses and plant circoviruses."
RL ARCH. VIROL. 143:1723-1744(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA NIAGRO F.D., FORSTHOEFFEL A.N., LAWTHOR R.P., KAMALANATHAN L.,
RITCHIE B.W., LATIMER K.S., LUKERT P.D.;
RL SUBMITTED (JUN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF071879; G3551845;
SQ SEQUENCE 230 AA; 27494 MW; 09F65071 CRC32;

Query Match 66.3%; Score 1196; DB 14; Length 230;
Best Local Similarity 66.5%; Pred. No. 1.65e-228;
Matches 153; Conservative 31; Mismatches 43; Indels 3; Gaps 2;

Db 1 MTWPRRRRRRTPRSHGLNLRPRPPLAHPAFNRVYRWREKGTGFSRLSTELVLRK 60
Qy 1 MTYPRRRRRRTPRSHGLNLRPRPPLAHPAFNRVYRWREKGTGFSRLSTELVLRK 58
Db 61 -GGYQSPPNNVYLFNIGCFPPSGGTPNPLPFPQYVIRKAKYEFYPRDPTISNQRGV 119
Qy 59 RTVTTPSWAVDMREKIDDFVPPGGGKNKISPEYVIRKVKVFWPCSPITQDGRV 118
Db 120 GSTVVILDANFVPTNLAIDPYINYSRHTIRQFTVHSRYFTPKPELDQIDWFHPNN 179
Qy 119 GSTVILDDNFVTKATALYDYPVYNSRHTIPQFYSHSRYFTPKPLDSTIDYFQNN 178
Db 180 KRNQLHLNHTVNEHTGLGVALONATAQNVYVRLIYVQFRELKD 229
Qy 179 KRNQLWLRLQSGNVHDVHGLGTAFENSKYDQDYNIRVTMYVQFRELKD 228

RESULT 6 PRELIMINARY; PRT; 244 AA.
ID O90237
AC O90237;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE CAPSID PROTEIN (FRAGMENT).
GN C1.
OS BEAK AND FEATHER DISEASE VIRUS.
OC VIRUSES; SSDNA VIRUSES; CIRCOVIRIDAE; CIRCOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 99003925.
RA NIAGRO F.D., FORSTHOEFFEL A.N., LAWTHOR R.P., KAMALANATHAN L.,
RITCHIE B.W., LATIMER K.S., LUKERT P.D.;
RT "Beak and feather disease virus and porcine circovirus genomes:
intermediates between the geminiviruses and plant circoviruses."
RL ARCH. VIROL. 143:1723-1744(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA NIAGRO F.D., FORSTHOEFFEL A.N., LAWTHOR R.P., KAMALANATHAN L.,
RITCHIE B.W., LATIMER K.S., LUKERT P.D.;
RL SUBMITTED (JUN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF071878; G3551842;
FT NON_TER
SQ SEQUENCE 244 AA; 28685 MW; 0164680E CRC32;

Query Match 12.9%; Score 233; DB 14; Length 244;
Best Local Similarity 31.7%; Pred. No. 2.59e-23;

```

```

atches. 77; Conservative 49; Mismatches 90; Indels 27; Gaps 21;

Db 11 RRRYARPYRRR-HI-RRYRRRRR--HFRRR-RFS-TNRVYTLRLRQFKFQIKQTST 64
Qy 5 RRRYRRRRPRSHGLQILRRRPWLHPRHRYRWKNGINFNRLSRFTGYTV-KRTT 63
Db 65 GNLISSDVFVTFALSDFIT--NTPNPTLNFENYRIKLAKMEMRPTWGHYTIQGD-GFGH 121
Qy 64 TP-SWAVDMREKIDDFVPPGGGKNKISPEYVIRKVKVFWPC-SPIT-QGDRGVGS 120
Db 122 TAVIODSRISKEFTVADQDQDPLAPDGARKVTVSRGK-RLLRPKPQITINDLTANQS 180
Qy 121 TAVILDDNVT-KATA-LTYDIPVYNSRHTIPQFYSHSRYFTPRP-V-LDSTIDYFQP 176
Db 181 AALWLNARTGWIPLQGGPNAAGAKVRHYGIGFSFPQPEQTIYVTKLTIYVQFROPAPN 240
Qy 177 NKR-N-QL-WLRQ----TSG-NVDHVLGTAFENSKYDQDYNIRVTMYVQFRELK 227
Db 241 NPS 243
Qy 228 DPP 230

RESULT 7 PRELIMINARY; PRT; 770 AA.
ID O70800
AC O70800;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE LONG ORF.
OS TT VIRUS.
OC VIRUSES; SSDNA VIRUSES; UNCLASSIFIED SSDNA VIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-G104901;
RA TAKAHASHI K., OHTA Y., MISHIRO S.;
RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AB011489; D1029096;
SQ SEQUENCE 770 AA; 90135 MW; EC5EEB7 CRC32;

Query Match 8.1%; Score 146; DB 14; Length 770;
Best Local Similarity 32.1%; Pred. No. 6.62e-08;
Matches 26; Conservative 17; Mismatches 34; Indels 4; Gaps 4;

Db 10 RRRWRWRPRWRRTTTRRRRPPARRRRNRVRRRGWRVRRYRWRKRRKRRK 69
Qy 5 RRRYRRRRPRSHGLQILRRRPWLHPRHRY-RWRKNGINFNRLSR-TF-GYTVKRTT 61
Db 70 IIRQWQPNYRRRCNIVGYIP 90
Qy 62 VTTPSWAVD-WMRFKIDDFVP 81

RESULT 8 PRELIMINARY; PRT; 770 AA.
ID O70796
AC O70796;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DE LONG ORF.
OS TT VIRUS.
OC VIRUSES; SSDNA VIRUSES; UNCLASSIFIED SSDNA VIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-G103301;
RA TAKAHASHI K., OHTA Y., MISHIRO S.;
RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AB011487; D1029092;
SQ SEQUENCE 770 AA; 90190 MW; A17FA5A9 CRC32;

Query Match 7.9%; Score 143; DB 14; Length 770;
Best Local Similarity 32.1%; Pred. No. 2.03e-07;
Matches 26; Conservative 16; Mismatches 35; Indels 4; Gaps 4;

```

```

Db 10 RRRWRRRRRPWRRTTTRRRPARRRRRRNRVRRRGGRWRRRRYRRWKKGRKKAK 69
||| ||| ||| : : ||| : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 5 RRRYRRRRHRRSHLQILRRRPLVHPRHRY-RWRRKNGIFNTRLSR-TF-GYTVKRTT 61
DE ORF1 PROTEIN.

Db 70 IIRWQPNYRRRCNIVGYLP 90
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 62 VTTPSWAVD-MMRFKIDDFVP 81

RESULT 9
ID O70739 PRELIMINARY; PRT; 770 AA.
AC O70739;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DE ORF1.
OS TT VIRUS.
OC VIRUSES; SSDNA VIRUSES; UNCLASSIFIED SSDNA VIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RA OKAMOTO H., NISHIZAWA T., KATO N., UKITA M., IKEDA H., IIZUKA H.,
RC STRAIN-TA278;
RA MIYAKAWA Y., MAYUMI M.;
RL HEPATOL. RES. 10-1-16(1998).
DR ENBL: AB008394; D1026059;
SQ SEQUENCE 770 AA; 90214 MW; 47BAB2F5 CRC32;

Query Match 7.9%; Score 143; DB 14; Length 770;
Best Local Similarity 32.1%; Pred. No. 2.03e-07;
Matches 26; Conservative 16; Mismatches 35; Indels 4; Gaps 4;

Db 10 RRRWRRRRRPWRRTTTRRRPARRRRRRNRVRRRGGRWRRRRYRRWKKGRKKAK 69
||| ||| ||| : : ||| : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 5 RRRYRRRRHRRSHLQILRRRPLVHPRHRY-RWRRKNGIFNTRLSR-TF-GYTVKRTT 61
DE ORF1 PROTEIN.

Db 70 IIRWQPNYRRRCNIVGYLP 90
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 62 VTTPSWAVD-MMRFKIDDFVP 81

RESULT 10
ID O70810 PRELIMINARY; PRT; 768 AA.
AC O70810;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DE LONG ORF.
OS TT VIRUS.
OC VIRUSES; SSDNA VIRUSES; UNCLASSIFIED SSDNA VIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-TY96117;
RA TAKAHASHI K., OHTA Y., MISHIRO S.;
RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR ENBL: AB011494; D1029106;
SQ SEQUENCE 768 AA; 90101 MW; 1FD49510 CRC32;

Query Match 7.8%; Score 140; DB 14; Length 768;
Best Local Similarity 31.3%; Pred. No. 6.17e-07;
Matches 25; Conservative 16; Mismatches 36; Indels 3; Gaps 3;

Db 10 RRRWRRRRRPWRRTTTRRRPARRRRRRNRVRRRGGRWRRRRYRRWKKGRKKAKI 69
||| ||| ||| : : ||| : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 5 RRRYRRRRHRRSHLQILRRRPLVHPRHRY-RWRRKNGIFNTRLSR-TF-GYTVKRTT 62
DE LONG ORF.

Db 70 IIRWQPNYRRRCNIVGYLP 89
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 63 TTPSWAVD-MMRFKIDDFVP 81

RESULT 11
ID O90364 PRELIMINARY; PRT; 770 AA.

```

```

AC O90364;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE ORF1 PROTEIN.
OS TT VIRUS.
OC VIRUSES; SSDNA VIRUSES; UNCLASSIFIED SSDNA VIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-TTVCHN1;
RA YUSEN Z., ZHONGPING H., YIHUA Z., HAITAO W.;
RT "Cloning and sequencing of TT virus from an Chinese patient with
hepatitis."
RL SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR ENBL: AF079173; G3378185;
SQ SEQUENCE 770 AA; 89881 MW; E5D48598 CRC32;

Query Match 7.7%; Score 139; DB 14; Length 770;
Best Local Similarity 50.0%; Pred. No. 8.91e-07;
Matches 20; Conservative 7; Mismatches 10; Indels 3; Gaps 3;

```

```

Db 23 PRRWTRRRPARRRGHRRNV-RRRRRGGRWRRRRYRRWKKR 61
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 4 PRRYRRRRH-RPRSHLQILRRRPLVHPRHRYR-WKKR 41
DE LONG ORF.

RESULT 12
ID O70804 PRELIMINARY; PRT; 769 AA.
AC O70804;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DE LONG ORF.
OS TT VIRUS.
OC VIRUSES; SSDNA VIRUSES; UNCLASSIFIED SSDNA VIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-G88801;
RA TAKAHASHI K., OHTA Y., MISHIRO S.;
RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR ENBL: AB011491; D1029100;
SQ SEQUENCE 769 AA; 90156 MW; 997B755A CRC32;

```

```

Query Match 7.4%; Score 133; DB 14; Length 769;
Best Local Similarity 30.0%; Pred. No. 7.91e-06;
Matches 24; Conservative 17; Mismatches 36; Indels 3; Gaps 3;

Db 10 RRRWRRRRRPWRRTTTRRRPARRRRRRNRVRRRGGRWRRRRYRRWKKGRKKAKI 69
||| ||| ||| : : ||| : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 5 RRRYRRRRHRRSHLQILRRRPLVHPRHRYR-WKKRNGIFNTRLSR-TF-GYTVKRTT 62
DE LONG ORF.

Db 70 IIRWQPNYRRRCNIVGYLP 89
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 63 TTPSWAVD-MMRFKIDDFVP 81

RESULT 13
ID O70802 PRELIMINARY; PRT; 769 AA.
AC O70802;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DE LONG ORF.
OS TT VIRUS.
OC VIRUSES; SSDNA VIRUSES; UNCLASSIFIED SSDNA VIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-G105001;
RA TAKAHASHI K., OHTA Y., MISHIRO S.;
RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR ENBL: AB011490; D1029098;
SQ SEQUENCE 769 AA; 90200 MW; 1366433E CRC32;

```





/db\_xref="GI:3668365"  
/translation="MVLLNCCPSRSLSEDFLGLILECYKTIHGGRGVVRHPPPM  
P"

## CDS

1016..1177

/codon\_start=1  
/product="P6.2"  
/protein\_id="AAC61863.1"  
/db\_xref="PID:3668366"  
/db\_xref="GI:3668366"  
/translation="MVFIIHLGFKWGVFKIKFSELYIHGYTDIVLVVFTVPSAEAE  
YVHHISRL"

## CDS

complement(1034..1735)

/codon\_start=1  
/product="P27.9"  
/protein\_id="AAC61864.1"  
/db\_xref="PID:3668367"  
/db\_xref="GI:3668367"  
/translation="MTYPRRRYRRRHRPSHLGQILRRPWLVPVHRRYRRKNGI  
FNRLSRFTGYTVKATVTTPSWAVDMRFKIDDFVPPGGTGNKISIPFEYVIRKVK  
VEFWPCSPITQDGRGVSTAVILDDNFVKATALTVDVYVNYSSRHTIPOPFYSHSRY  
ETPKPVLDSITDYFQPNKRNQLWLRLQTSNGVDHVLGTAFENSKYDODYINIRVTMY  
VQREFNLKDPLEP"

BASE COUNT  
ORIGIN

463 a 362 c 481 g 462 t

## alignment\_scores:

Quality: 318.00 Length: 59  
Ratio: 5.390 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-209-961-20 x AF085695 ..

Align seg 1/1 to: AF085695 from: 1 to: 1768

1 MetTyrThrSerLeuTTPGlyHisLeuGlyValValLysAlaAsnGlyLe 17  
|||||  
553 ATGTACACTTCATTTGGGGCCACCTGGTGTGTAAGCAATGGGCT 602

17 uLeuileLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrS 34

|||||

603 GCTAATTTTCAAAACCGGAAACACACATCTGGAACACCACTAAACAA 652

34 erGlyGlyMetValThrMetValLysLysTrpLeuLeuMetThrPhe 50

|||||

653 GTGGTGGGATGGTTACCATGGTGAAAGTGGTGTATTGATGACTTTT 702

51 MetAlaGlyCysArgGlyMetIleTyr 59

|||||

703 ATGGCTGGCTGCCGTGGGATGATCTAC 729

seq\_name: gb\_vi:AF086834

## seq\_documentation\_block:

LOCUS AF086834 1768 bp DNA circular VRL 29-SEP-1998

DEFINITION Porcine circovirus strain B9, complete genome.

ACCESSION AF086834

NID 93661515

VERSION AF086834.1 GI:3661515

KEYWORDS porcine circovirus.

SOURCE porcine circovirus

ORGANISM Viruses; ssDNA viruses; Circoviridae; Circovirus.

REFERENCE 1 (bases 1 to 1768)

AUTHORS Wang,L., Willson,P., Chow,B., Gibbons,E. and Babiuk,L.

TITLE Emergence of a new porcine circovirus

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1768)

AUTHORS Wang,L., Willson,P., Chow,B., Gibbons,E. and Babiuk,L.

TITLE Direct Submission

JOURNAL Submitted (26-AUG-1998) VIDO, University of Saskatchewan, 120

Veterinary Road, Saskatoon, Saskatchewan S7N 5E3, Canada

FEATURES  
source

1..1768  
/organism="porcine circovirus"  
/strain="B9"  
/db\_xref="taxon:46221"  
51..995

## CDS

/codon\_start=1  
/product="P35.8"  
/protein\_id="AAC61737.1"  
/db\_xref="PID:93661516"  
/db\_xref="GI:3661516"  
/translation="MPSKNGRSGQPQHKRWFTLNNPSEDEKKIKELPISLFDYFI  
VGEENEGRTPHLOGFANFVKQTKNFKVLCGARCHIEKAGTDOONKEYCKSGKM  
LLIECGAPRSQORSLSLAVSTLLESGILVTAKOHPVTFVKNFRGLAELLKYSKGM  
OKRDKNTYHFIVGPPCGCKSKWAANFETTYWPKPKNWDGIGEKVWVDDYF  
GWLDDLLRLCDRIEFTVTKGTGTFPLARSILITSNQTPLEWISSSTAVFAVEALYR  
RTSLVFNKNTEQSQEGGQFVTLSPPCPEFFEYIN"

## CDS

complement(1034..1735)  
/codon\_start=1  
/product="P27.9"  
/protein\_id="AAC61738.1"  
/db\_xref="PID:93661517"  
/db\_xref="GI:3661517"  
/translation="MTYPRRRYRRRHRPSHLGQILRRPWLVPVHRRYRRKNGI  
FNRLSRFTGYTVKATVTTPSWAVDMRFNIDDFVPPGGTGNKISIPFEYVIRKVK  
VEFWPCSPITQDGRGVSTAVILDDNFVKATALTVDVYVNYSSRHTIPOPFYSHSRY  
ETPKPVLDSITDYFQPNKRNQLWLRLQTSNGVDHVLGTAFENSKYDQDYNIRVTMY  
VQREFNLKDPPLKP"

BASE COUNT 463 a 357 c 482 g 466 t  
ORIGIN

## alignment\_scores:

Quality: 318.00 Length: 59  
Ratio: 5.390 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-209-961-20 x AF086834 ..

Align seg 1/1 to: AF086834 from: 1 to: 1768

1 MetTyrThrSerLeuTTPGlyHisLeuGlyValValLysAlaAsnGlyLe 17  
|||||  
553 ATGTACACTTCATTTGGGGCCACCTGGTGTGTAAGCAATGGGCT 602

17 uLeuileLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrS 34

|||||

603 GCTAATTTTCAAAACCGGAAACACACATCTGGAACACCACTAAACAA 652

34 erGlyGlyMetValThrMetValLysLysTrpLeuLeuMetThrPhe 50

|||||

653 GTGGTGGGATGGTTACCATGGTGAAAGTGGTGTATTGATGACTTTT 702

51 MetAlaGlyCysArgGlyMetIleTyr 59

|||||

703 ATGGCTGGCTGCCGTGGGATGATCTAC 729

seq\_name: gb\_vi:AF086835

## seq\_documentation\_block:

LOCUS AF086835 1768 bp DNA circular VRL 29-SEP-1998

DEFINITION Porcine circovirus strain 9741, complete genome.

ACCESSION AF086835

NID 93661518

VERSION AF086835.1 GI:3661518

KEYWORDS porcine circovirus.

SOURCE porcine circovirus

ORGANISM Viruses; ssDNA viruses; Circoviridae; Circovirus.

REFERENCE 1 (bases 1 to 1768)

AUTHORS Wang,L., Willson,P., Chow,B., Gibbons,E. and Babiuk,L.

TITLE Emergence of a new porcine circovirus

JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1768)  
AUTHORS Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.  
TITLE Direct Submission  
JOURNAL Submitted (26-AUG-1998) VIDO, University of Saskatchewan, 120  
Veterinary Road, Saskatoon, Saskatchewan S7N 5E3, Canada  
FEATURES  
Location/Qualifiers  
source  
1. .1768  
/organism="porcine circovirus"  
/strain="9741"  
/db\_xref="taxon:46221"  
51. .995  
/codon\_start=1  
/product="p35.8"  
/protein\_id="AAC61739.1"  
/db\_xref="PID:g3661519"  
/db\_xref="GI:3661519"  
CDS  
/translation="MPSKNGRSGPQPHKRWFTLNNPSENKRRKKIRELPISLFDYFI  
VGEENEGRTPLHGFQFANFKVKTENKVKWYLGARCHIEKAKGTDOONKEYCSKGN  
LLIECGAPRSGORSLSLSTAVSTLLESGILTVAKQHPVTVKNGRLAELLKVSCKM  
OKRDWNTNHFIVGPPCGSKKWAANFANPETTYWPKPKNKWDGYHGEKVVVDIFY  
GWLPPDLDRLCDRIPLTVKGTGTFPLARSILITSNQTPLEWYSSTAVPAVEALIR  
RITSLVFNKATQSTEEGQFVTLSPPCPEFFEYIN"  
complement(1034. .1735)  
/codon\_start=1  
/product="p27.9"  
/protein\_id="AAC61740.1"  
/db\_xref="PID:g3661520"  
/db\_xref="GI:3661520"  
CDS  
/translation="MTYPRRRYRRRRHRRSHLGOILRRRPLVHPHRRYWRKNGI  
FNRLSRTFGYTVKRTVTTPSWAVDMRKFIDDFVPPGGTGNKISIPFEYIRKVK  
VEFWPCSPITQDRCGVSTAVILDNFTKATALTYPYNYSSRHTIPQPSYHSRY  
FTPKPVLDSIDYFQPNKRNQLWRLQTSNVDHVLGTAFENSKYDQDYNIRVTY  
VQPREFNKDPLEP"  
464 a 360 c 480 g 464 t  
BASE COUNT  
ORIGIN  
alignment\_scores:  
Quality: 318.00 Length: 59  
Ratio: 5.390 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-09-209-961-20 x AF086835 ..  
Align seg 1/1 to: AF086835 from: 1 to: 1768  
1 MetTyrThrSerLeuTrpGlyHisLeuGlyValValLysAlaAsnGlyLe 17  
553 ATGTACACTTCATTGGGGCCACCTGGTGTGTAAGCAAAATGGGCT 602  
17 uLeuileLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrS 34  
603 GCTAATTTGCAACCCGGAACCACTACTGGAACCACTTAAACAA 652  
34 erGlyGlyMetValThrMetValLysLysTrpLeuLeuLeuMetThrPhe 50  
653 GTGGTGGGATGTTACCATGGTGAAGAGTGTGTTATTGATGACATTT 702  
51 MetAlaGlyCysArgGlyMetIleTyr 59  
703 ATGGCTGGCTGCCGTGGGATGATCTAC 729  
seq\_name: gb\_vi:AF086836  
seq\_documentation\_block:  
LOCUS AF086836 1768 bp DNA circular VRL 29-SEP-1998  
DEFINITION Porcine circovirus strain M226, complete genome.  
ACCESSION AF086836  
NID g3661521  
VERSION AF086836.1 GI:3661521  
KEYWORDS

SOURCE  
ORGANISM  
REFERENCE 1 (bases 1 to 1768)  
AUTHORS Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.  
TITLE Emergence of a new porcine circovirus  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1768)  
AUTHORS Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.  
TITLE Direct Submission  
JOURNAL Submitted (26-AUG-1998) VIDO, University of Saskatchewan, 120  
Veterinary Road, Saskatoon, Saskatchewan S7N 5E3, Canada  
FEATURES  
Location/Qualifiers  
source  
1. .1768  
/organism="porcine circovirus"  
/strain="M226"  
/db\_xref="taxon:46221"  
51. .995  
/codon\_start=1  
/product="p35.8"  
/protein\_id="AAC61741.1"  
/db\_xref="PID:g3661522"  
/db\_xref="GI:3661522"  
CDS  
/translation="MPSKNGRSGPQPHKRWFTLNNPSEDERKKIRELPISLFDYFI  
VGEENEGRTPLHGFQFANFKVKTENKVKWYLGARCHIEKAKGTDOONKEYCSKGN  
LLIECGAPRSGORSLSLSTAVSTLLESGILTVAKQHPVTVKNGRLAELLKVSCKM  
OKRDWNTNHFIVGPPCGSKKWAANFANPETTYWPKPKNKWDGYHGEKVVVDIFY  
GWLPPDLDRLCDRIPLTVKGTGTFPLARSILITSNQTPLEWYSSTAVPAVEALIR  
RITSLVFNKATQSTEEGQFVTLSPPCPEFFEYIN"  
complement(1034. .1735)  
/codon\_start=1  
/product="p27.9"  
/protein\_id="AAC61742.1"  
/db\_xref="PID:g3661523"  
/db\_xref="GI:3661523"  
CDS  
/translation="MTYPRRRYRRRRHRRSHLGOILRRRPLVHPHRRYWRKNGI  
FNRLSRTFGYTVKRTVTTPSWAVDMRKFIDDFVPPGGTGNKISIPFEYIRKVK  
VEFWPCSPITQDRCGVSTAVILDNFTKATALTYPYNYSSRHTIPQPSYHSRY  
FTPKPVLDSIDYFQPNKRNQLWRLQTSNVDHVLGTAFENSKYDQDYNIRVTY  
VQPREFNKDPLEP"  
463 a 363 c 480 g 462 t  
BASE COUNT  
ORIGIN  
alignment\_scores:  
Quality: 318.00 Length: 59  
Ratio: 5.390 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-09-209-961-20 x AF086836 ..  
Align seg 1/1 to: AF086836 from: 1 to: 1768  
1 MetTyrThrSerLeuTrpGlyHisLeuGlyValValLysAlaAsnGlyLe 17  
553 ATGTACACTTCATTGGGGCCACCTGGTGTGTAAGCAAAATGGGCT 602  
17 uLeuileLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrS 34  
603 GCTAATTTGCAACCCGGAACCACTACTGGAACCACTTAAACAA 652  
34 erGlyGlyMetValThrMetValLysLysTrpLeuLeuLeuMetThrPhe 50  
653 GTGGTGGGATGTTACCATGGTGAAGAGTGTGTTATTGATGACATTT 702  
51 MetAlaGlyCysArgGlyMetIleTyr 59  
703 ATGGCTGGCTGCCGTGGGATGATCTAC 729  
seq\_name: gb\_vi:AF027217  
seq\_documentation\_block:

```

LOCUS       AF027217      1768 bp      DNA      circular      VRL      14-MAY-1998
DEFINITION  Porcine circovirus strain pmws PCV, complete genome.
ACCESSION   AF027217
NID         92689645
VERSION     AF027217.1  GI:2689645
SOURCE      porcine circovirus.
  ORGANISM  Viruses: ssDNA viruses; Circoviridae; Circovirus.
REFERENCE   1 (bases 1 to 1768)
  AUTHORS   Hamel,A.L., Lin,L.L. and Nayar,G.P.
  TITLE     Nucleotide sequence of porcine circovirus associated with
  JOURNAL   postweaning multisystemic wasting syndrome in pigs
  MEDLINE   J. Virol. 72 (6), 5262-5267 (1998)
  ID        98241772
REFERENCE   2 (bases 1 to 1768)
  AUTHORS   Hamel,A.L., Lin,L.L. and Nayar,G.P.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-SEP-1997) Virology Laboratory, Veterinary Services
  ID        Branch, Manitoba Agriculture, 545 University Crescent, Winnipeg,
  ID        Manitoba R3T 5S6, Canada
FEATURES             Location/Qualifiers
     source          1..1768
                     /organism="porcine circovirus"
                     /strain="pmws PCV"
                     /db_xref="taxon:46221"
                     /note="both strands of seven overlapping PCR fragments
                     were sequenced: virus isolated from lung, lymph node,
                     spleen and tonsil tissue from pigs affected by post
                     weaning multisystemic wasting syndrome"
                     51..995
     CDS             1..995
                     /note="ORF1; similar to Rep protein encoded by
                     non-pathogenic PCV, GenBank Accession Number U49186;
                     predicted 35.8 kDa protein"
                     /codon_start=1
                     /product="putative Rep protein"
                     /protein_id="AAC59462.1"
                     /db_xref="PID:g2689646"
                     /db_xref="GI:2689646"
                     /translation="MPSKNGRSGQPQHRVWFTLNPNSEDERKKIRELPISLFDFYI
                     VEGEGNEGRTPHLOGFANFKQTFNKVKYLGARCIERKARGDQQNKYCSKEGN
                     LLIEGAPRSQSGDLSSTAVSLLESGLTVAEQHPVTFVNRFRGLAELLKVGSKM
                     OKRDKNTNVHTVGGGCKSKAAAFADPTTYNKPKNKWDGYPHGEVIVIDFY
                     GWLPWDLRLCDRLPLVTETKGGTVPLARSILITSNOTPLEWTSSTAPVAEALYR
                     RTLSLVFNKNATEQSTEEGGQFTLSPPCPPEYIN"
                     117..125
     misc_feature    117..125
                     /note="glycosylation site"
     polyA_signal    327..332
     CDS             complement(357..671)
                     /note="ORF3; predicted 11.9 kDa protein"
                     /codon_start=1
                     /protein_id="AAC59464.1"
                     /db_xref="PID:g2689648"
                     /db_xref="GI:2689648"
                     /translation="MVTIPPLVSRWFPVCGFRVKISSPFAFTTPRPHNDVYIGLPI
                     TLLHPFAHFQKFSQPAEISDKRYVLLCNGHOTPALQOQTHSSRQVTPLSLSRSSTF
                     NK"
     CDS             complement(386..565)
                     /note="ORF4; predicted 6.5 kDa protein"
                     /codon_start=1
                     /protein_id="AAC59465.1"
                     /db_xref="PID:g2689649"
                     /db_xref="GI:2689649"
                     /translation="MTCVLVFSQRCIFPLTFKSSASPRKFLTNVTGCCSATVTRLPL
                     SNKVLTAVDRSLRUC"
     misc_feature    complement(470..478)
     CDS             complement(688..753)
                     /note="glycosylation site"
                     /codon_start=1
                     /note="ORF8; predicted 2.3 kDa protein"
                     /protein_id="AAC59469.1"
                     /db_xref="PID:g2689653"
                     /db_xref="GI:2689653"

```

```

/translation="MDIDHTVSDVHPTAAASHKSHQ"
816..824
/note="glycosylation site"
906..914
/note="glycosylation site"
983..988
complement(989..1033)
/note="ORF1; predicted 1.8 kDa protein"
/codon_start=1
/protein_id="AAC59472.1"
/db_xref="PID:g2689656"
/db_xref="GI:2689656"
/translation="MKNNHYEVIKKTQ"
1016..1177
/note="ORF5; predicted 6.2 kDa protein"
/codon_start=1
/protein_id="AAC59466.1"
/db_xref="PID:g2689650"
/db_xref="GI:2689650"
/translation="WVFIFHLGFKMGVFKIKFSELYIHGYTDIWLVLVYTVFERSAEA
YVHISRG"
complement(1022..1027)
complement(1034..1735)
/note="ORF2; predicted 27.8 kDa protein"
/codon_start=1
/protein_id="AAC59463.1"
/db_xref="PID:g2689647"
/db_xref="GI:2689647"
/translation="MTYPRRYRRRRPFRSHGLQILRRPWLPHRHYRWRKNGKI
FNTRLSRTFGYTKATVTPSWAVDMRFNIDDFVPGGNTKISIPFYIRKVKY
VEFWPCSPITQDGRGVSTAVILDDNFVKATALTDPYVYNSRHTIPQPFYHSRY
FVFPKPLDSTIDYFQPNKRTQLWLRLQTSRNVHDVHGLGTAFAENSIYDQDYNIRVTWY
VQFREFNLKDPPLKP"
complement(1301..1309)
/note="glycosylation site"
complement(1522..1611)
/note="ORF6; predicted 3.1 kDa protein"
/codon_start=1
/protein_id="AAC59467.1"
/db_xref="PID:g2689651"
/db_xref="GI:2689651"
/translation="MASTPASPAPSDILSRLPQSRPPGRWT"
1524..1631
/note="ORF10; predicted 4.1 kDa protein"
/codon_start=1
/protein_id="AAC59471.1"
/db_xref="PID:g2689655"
/db_xref="GI:2689655"
/translation="MSTAQEGVLTVALTVYKVRERRVLKMPFFLIQR"
1682..1741
/note="ORF7; predicted 1.9 kDa protein"
/codon_start=1
/protein_id="AAC59468.1"
/db_xref="PID:g2689652"
/db_xref="GI:2689652"
/translation="MAAGAVSSSAVTPPWIRHS"
complement(1732..1768,1..92))
/note="ORF9; predicted 4.6 kDa protein"
/codon_start=1
/protein_id="AAC59470.1"
/db_xref="PID:g2689654"
/db_xref="GI:2689654"
/translation="MWLGASILLAGHVAEVLPRCCRCRSALVILTAHFFRQL"
join(1746..1768,1..13)
/function="putative replication site"
join(1762..1768,1..2)
/note="'AAGTATTAC' is similar to the nonanucleotide motif
in the non-pathogenic PCV, GenBank Accession Number
U49186"

```

BASE COUNT 452 a 360 c 495 g 461 t  
ORIGIN

```
alignment_scores:
  Quality: 314.00      Length: 59
  Ratio: 5.322        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 98.305

alignment_block:
  US-09-209-961-20 x AF027217 ..

  Align seg 1/1 to: AF027217 from: 1 to: 1768

    1 MetTyrThrSerLeuTrpGlyHisLeuGlyValValLysAlaAsnGlyLe 17
      |||||
    553 ATGTACAGCTCATTTGGGGCCACCTGGTGTGTAAGAAAGCAATGGCT 602
      |||||

    17 uLeuLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrS 34
      |||||
    603 GCTAATTTGAGACCCGGAACACACATACATGGAACACCATAGAACAA 652
      |||||

    34 erGlyGlyMetValThrMetValLysLysTyrLeuLeuLeuMetThrPhe 50
      |||||
    653 GTGGTGGATGTTACCATGCTGAAGAGTGGTGTATTGATGACTTTT 702
      |||||

    51 MetAlaGlyCysArgGlyMetIleTyr 59
      |||||
    703 ATGGCTGGTGGCGGTGGATGATCTAC 729

  seq_name: gb_vi:AF055391

  seq_documentation_block:
  LOCUS AF055391 1768 bp DNA circular VRL 13-SEP-1998
  DEFINITION Porcine circovirus Type II from USA, complete genome.
  ACCESSION AF055391
  NID 93598796
  VERSION AF055391.1 GI:3598796
  SOURCE
  ORGANISM
  porcine circovirus.
  porcine circovirus
  Viruses; ssDNA viruses; Circoviridae; Circovirus.
  REFERENCE 1 (bases 1 to 1768)
  Meehan,B.M., McNeilly,F., Todd,D., Kennedy,S., Jewhurst,V.A.,
  Ellis,J.A., Hassard,L.E., Clark,E.G., Haines,D.M. and Allan,G.M.
  Characterization of novel circovirus DNAs associated with wasting
  syndromes in pigs
  J. Gen. Virol. 79 (Pt 9), 2171-2179 (1998)
  MEDLINE 98418498
  REFERENCE 2 (bases 1 to 1768)
  Meehan,B.M.
  Direct Submission
  Submitted (26-MAR-1998) Brian M Meehan, Veterinary Science, The
  Queen's University of, Belfast, Stoney Road, Belfast, Northern
  Ireland, BT4 3SD, U.K.
  FEATURES
  Location/Qualifiers
    1..1768
      /organism="porcine circovirus"
      /specific_host="Sus scrofa"
      /strain="Porcine circovirus Type II"
      /db_xref="taxon:46221"
      /clone="Imp.999 ECORI No.8"
      /country="USA"
      complement(101..190)
      /codon_start=1
      /product="ORF6"
      /protein_id="AAC35308.1"
      /db_xref="PID:g3598807"
      /db_xref="GI:3598807"
      /translation="MASSTPASPAPSDILSRLPQSRPPGRWT"
      103..210
      /codon_start=1
      /product="ORF10"
      /protein_id="AAC35306.1"
      /db_xref="PID:g3598805"
      /db_xref="GI:3598805"
      /translation="MSTAQEGVLTVVATVYPKVRERRVLKMPFFLLQR"

  CDS
    261..320
      /codon_start=1
      /product="ORF7"
      /protein_id="AAC35307.1"
      /db_xref="PID:g3598806"
      /db_xref="GI:3598806"
      /translation="MAAGAVSSAVTPPWIRHS"
      complement(311..439)
      /note="predicted 4.6 kDa protein"
      /codon_start=1
      /product="ORF9"
      /protein_id="AAC35304.1"
      /db_xref="PID:g3598803"
      /db_xref="GI:3598803"
      /translation="MWLGSSASSILLAGHVAEVLPRCCRCRSALVILTAHFRFQL"
      398..1342
      /codon_start=1
      /product="putative Rep protein (ORF1)"
      /protein_id="AAC35298.1"
      /db_xref="PID:g3598797"
      /db_xref="GI:3598797"
      /translation="MPSKKNGRSGPQPHKRWFTLNNPSEDERKKIRELPISLDFYFI
      VGEENEGRTPLHQLGFANFYKOTFNKRWYLGARCHIEKAGTDQONKEYCSKEGN
      LLIECGAPRSGQSRSDLSSTAVSTLLESGSLVVAEQHPVTFYRNRGLAEVLKSGKM
      QKRWKTNVIVGPPGCGSKWAANFADPTTYWKPNNKMWDCYHGEEVVIIDFY
      GWLPWDDLRLCDRYPLTVETKGTVPFLARSILITISNQTPLEWYSTTAVPAVEALYR
      RITSLVFNKNAEQSTEEGGFVTLSPPCPEFYNT"
      674..679
      complement(704..1018)
      /note="predicted 11.9 kDa protein"
      /codon_start=1
      /product="ORF3"
      /protein_id="AAC35300.1"
      /db_xref="PID:g3598799"
      /db_xref="GI:3598799"
      /translation="MVTIPPLVSRWFPVCGFRVCKISSPFAFTTTPRWPHNDVYIGLPI
      TLLHFAHFOKFSQPAEISDKRYRLLCNGHQTALQOQTHSSRQVTPILSRSSSTF
      NK"
      complement(733..912)
      /note="predicted 6.5 kDa protein"
      /codon_start=1
      /product="ORF4"
      /protein_id="AAC35301.1"
      /db_xref="PID:g3598800"
      /db_xref="GI:3598800"
      /translation="MTCTLVQSRFCIFPLTFKSSASPRKFLTNVTCGSATVTRLPL
      SNKVLTAVDRLSRCP"
      complement(1035..1100)
      /note="predicted 2.3 kDa protein"
      /codon_start=1
      /product="ORF8"
      /protein_id="AAC35303.1"
      /db_xref="PID:g3598802"
      /db_xref="GI:3598802"
      /translation="MDIDHTVSDVHPTAAASHKSHQ"
      1330..1335
      complement(1336..1380)
      /note="predicted 1.8 kDa protein"
      /codon_start=1
      /product="ORF11"
      /protein_id="AAC35305.1"
      /db_xref="PID:g3598804"
      /db_xref="GI:3598804"
      /translation="MNNKNHYEVKKQTQ"
      1363..1524
      /note="predicted 6.2 kDa protein"
      /codon_start=1
      /product="ORF5"
      /protein_id="AAC35302.1"
      /db_xref="PID:g3598801"
      /db_xref="GI:3598801"
      /translation="MVFIIHLGFWGVKIKFSELYINGYTDIVVLVVTYVFSRSEA
      YVWHISRGL"

  CDS
    261..320
      /codon_start=1
      /product="ORF7"
      /protein_id="AAC35307.1"
      /db_xref="PID:g3598806"
      /db_xref="GI:3598806"
      /translation="MAAGAVSSAVTPPWIRHS"
      complement(311..439)
      /note="predicted 4.6 kDa protein"
      /codon_start=1
      /product="ORF9"
      /protein_id="AAC35304.1"
      /db_xref="PID:g3598803"
      /db_xref="GI:3598803"
      /translation="MWLGSSASSILLAGHVAEVLPRCCRCRSALVILTAHFRFQL"
      398..1342
      /codon_start=1
      /product="putative Rep protein (ORF1)"
      /protein_id="AAC35298.1"
      /db_xref="PID:g3598797"
      /db_xref="GI:3598797"
      /translation="MPSKKNGRSGPQPHKRWFTLNNPSEDERKKIRELPISLDFYFI
      VGEENEGRTPLHQLGFANFYKOTFNKRWYLGARCHIEKAGTDQONKEYCSKEGN
      LLIECGAPRSGQSRSDLSSTAVSTLLESGSLVVAEQHPVTFYRNRGLAEVLKSGKM
      QKRWKTNVIVGPPGCGSKWAANFADPTTYWKPNNKMWDCYHGEEVVIIDFY
      GWLPWDDLRLCDRYPLTVETKGTVPFLARSILITISNQTPLEWYSTTAVPAVEALYR
      RITSLVFNKNAEQSTEEGGFVTLSPPCPEFYNT"
      674..679
      complement(704..1018)
      /note="predicted 11.9 kDa protein"
      /codon_start=1
      /product="ORF3"
      /protein_id="AAC35300.1"
      /db_xref="PID:g3598799"
      /db_xref="GI:3598799"
      /translation="MVTIPPLVSRWFPVCGFRVCKISSPFAFTTTPRWPHNDVYIGLPI
      TLLHFAHFOKFSQPAEISDKRYRLLCNGHQTALQOQTHSSRQVTPILSRSSSTF
      NK"
      complement(733..912)
      /note="predicted 6.5 kDa protein"
      /codon_start=1
      /product="ORF4"
      /protein_id="AAC35301.1"
      /db_xref="PID:g3598800"
      /db_xref="GI:3598800"
      /translation="MTCTLVQSRFCIFPLTFKSSASPRKFLTNVTCGSATVTRLPL
      SNKVLTAVDRLSRCP"
      complement(1035..1100)
      /note="predicted 2.3 kDa protein"
      /codon_start=1
      /product="ORF8"
      /protein_id="AAC35303.1"
      /db_xref="PID:g3598802"
      /db_xref="GI:3598802"
      /translation="MDIDHTVSDVHPTAAASHKSHQ"
      1330..1335
      complement(1336..1380)
      /note="predicted 1.8 kDa protein"
      /codon_start=1
      /product="ORF11"
      /protein_id="AAC35305.1"
      /db_xref="PID:g3598804"
      /db_xref="GI:3598804"
      /translation="MNNKNHYEVKKQTQ"
      1363..1524
      /note="predicted 6.2 kDa protein"
      /codon_start=1
      /product="ORF5"
      /protein_id="AAC35302.1"
      /db_xref="PID:g3598801"
      /db_xref="GI:3598801"
      /translation="MVFIIHLGFWGVKIKFSELYINGYTDIVVLVVTYVFSRSEA
      YVWHISRGL"
```



```
polyA_site      complement(1369..1374)
CDS             complement(join(1381..1768,1..314))
               /note="predicted 27.8 kDa protein"
               /codon_start=1
               /product="ORF2"
               /protein_id="AAC35299.1"
               /db_xref="PID:g3598798"
               /db_xref="GI:3598798"
               /translation="MTYPRRRYRRRRHPRSHLGLILRRRPLVHPHRYRWRKNGI
ENTLSRTEGTVKATTVRTPSWAVDMRENIIDFVPGGNTKISIPPEYRIKKVK
VEFWPCSIQTGDRGVGSFAVILNDNFYTKALYIDPYVNSRHITPOPEYSISRY
ETPFPVLDSTIDFQPNKRNQLMLRLTSRNVHVLGLTAFENSIYDQYNIIRVTM
VOFRENLKDPKLP"
BASE COUNT     453 a 360 c 494 g 461 t
ORIGIN

alignment_scores:
  Quality: 314.00      Length: 59
  Ratio: 5.322        Gaps: 0
  Percent similarity: 100.000  Percent identity: 98.305

alignment_block:
US-09-209-961-20 x AF055391 ..
Align seg 1/1 to: AF055391 from: 1 to: 1768
1 MetTyrThrSerLeuTrpGlyHisLeuGlyValVallysalasnglyLe 17
|||||
900 ATGTACACGTCATGTGGGCCACCTGGTGTGTAAGCAATGGGCT 949
|||||
17 uLeulleLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrs 34
|||||
950 GCTAATTTTCAGACCCGGAACACACATACATGGAACACCTAGAACAA 999
|||||
34 erGlyGlyMetValThrMetValLysLysTrpLeuLeuMetThrPhe 50
|||||
1000 GTGGTGGGATGGTTACCATGGTGAAGAAGTGTGTATTGATGACTTTT 1049
|||||
51 MetAlaGlyCysArgGlyMetIleTyr 59
|||||
1050 ATGGCTGGCTGCGCGGTGGATGATCTAC 1076
|||||
seq_name: gb_vi:AF055392

seq_documentation_block:
LOCUS      AF055392      1768 bp      DNA      circular      VRL      13-SEP-1998
DEFINITION Porcine circovirus Type II from Canada, complete genome.
ACCESSION  AF055392
NID        g3598808
VERSION    AF055392.1 GI:3598808
KEYWORDS   porcine circovirus.
SOURCE     porcine circovirus
ORGANISM   Viruses; ssDNA viruses; Circoviridae; Circovirus.
REFERENCE  1 (bases 1 to 1768)
AUTHORS   Meehan,B.M., McNeill,F., Todd,D., Kennedy,S., Jewhurst,V.A.,
Ellis,J.A., Hassard,L.E., Clark,E.G., Haines,D.M. and Allan,G.M.
TITLE     Characterization of novel circovirus DNAs associated with wasting
syndromes in pigs
JOURNAL   J. Gen. Virol. 79 (Pt 9), 2171-2179 (1998)
MEDLINE   98418498
REFERENCE  2 (bases 1 to 1768)
AUTHORS   Meehan,B.M.
TITLE     Direct Submission
JOURNAL   Submitted (26-MAR-1998) Brian M Meehan, Veterinary Science, The
Queen's University of Belfast, Stoney Road, Belfast, Northern
Ireland, BT4 3SD, U.K.
FEATURES   Location/Qualifiers
            1..1768
            /organism="porcine circovirus"
            /specific_host="Sus scrofa"
            /strain="Porcine circovirus Type II"
```

```
/db_xref="taxon:46221"
/clone="Imp.1010-Stoon ECORI No.14"
/country="Canada"
complement(101..190)
/codon_start=1
/product="ORF6"
/protein_id="AAC35319.1"
/db_xref="PID:g3598819"
/db_xref="GI:3598819"
/translation="MASSTPASPAPSDILSSVPQSORPGRWT"
103..210
/codon_start=1
/product="ORF10"
/protein_id="AAC35317.1"
/db_xref="PID:g3598817"
/db_xref="GI:3598817"
/translation="MSTAGGVTVVRLTVYPKVRRRLVKMPFFLLQR"
261..320
/codon_start=1
/product="ORF7"
/protein_id="AAC35318.1"
/db_xref="PID:g3598818"
/db_xref="GI:3598818"
/translation="MAAGAVSSAVTPPWIRHS"
complement(311..439)
/note="predicted 4.6 kDa protein"
/codon_start=1
/product="ORF9"
/protein_id="AAC35315.1"
/db_xref="PID:g3598815"
/db_xref="GI:3598815"
/translation="MWLGSASSILLAGHVAEEVLPCCRCRSALVILTAHFFRFQL"
398..1342
/codon_start=1
/product="putative Rep protein (ORF1)"
/protein_id="AAC35309.1"
/db_xref="PID:g3598809"
/db_xref="GI:3598809"
/translation="MPSKNGSGPOPHKRWVTLNPNSEDEBKIRELPISLDFYFI
VGEENEGRTPHLQGFANFKVKNWLGARCHEKAGTKQOQNKKEYCSKGN
LLIEGAPRSQGRSDLSAVSLTLESGLTVAEQHPVTFVNFGLAELLKVSCKM
QKRWKTNVHVIPPGCKSKAANFADPETTYKPPRNKMWGDSYHGEVVYIDDFY
GWLWDDLLRLCDRYPLTVETKGGTVPLARSILITSNQTPLEWYSSTAVPAVEALYR
RITSLVFWKNATEQSTEEGQFVTLSPCCPEPYEINY"
674..679
complement(704..1018)
/note="predicted 11.9 kDa protein"
/codon_start=1
/product="ORF3"
/protein_id="AAC35311.1"
/db_xref="PID:g3598811"
/db_xref="GI:3598811"
/translation="MVTIPPLVSRMFPVCGFRVCKISSPFAFTTPRPHNDVYIGLPI
TLLHFPAPHFQESQPAEISDKRYRVLLCNGHOTPALQOQTHSSRQVTPLSLRSRSTF
NR"
complement(733..912)
/note="predicted 6.5 kDa protein"
/codon_start=1
/product="ORF4"
/protein_id="AAC35312.1"
/db_xref="PID:g3598812"
/db_xref="GI:3598812"
/translation="MTCILVFQSRFCIFPLTFKSSASPRKFLTNVTCGCSATVRLPL
SNKVLTAVDRLSRCP"
complement(1035..1100)
/note="predicted 2.3 kDa protein"
/codon_start=1
/product="ORF8"
/protein_id="AAC35314.1"
/db_xref="PID:g3598814"
/db_xref="GI:3598814"
/translation="MDIDHTVSDVHDPTAASHKSHQ"
1330..1335
polyA_signal
```

```

CDS
complement(1336..1380)
/note="predicted 1.8 kDa protein"
/codon_start=1
/product="ORF1"
/protein_id="AAC35316.1"
/db_xref="PID:g3598816"
/db_xref="GI:3598816"
/translation="MNNKNHYEVIKKTQ"
1363..1524
/note="predicted 6.2 kDa protein"
/codon_start=1
/product="ORF3"
/protein_id="AAC35313.1"
/db_xref="PID:g3598813"
/db_xref="GI:3598813"
/translation="MVFIIHLGFRWGVKIKFSELYIHGYTDIVLVVTFVFSRSEA
YVHISRGL"
polyA_site
complement(1369..1374)
complement(join(1381..1768,1..314))
/note="predicted 27.8 kDa protein"
/codon_start=1
/product="ORF2"
/protein_id="AAC35310.1"
/db_xref="PID:g3598810"
/db_xref="GI:3598810"
/translation="MTVPRRRYRRRRHPRSHLGOILRRPWLVPVPRHYRWRKNGI
FNFLSTFGYTKRTVTTPSWAVDMRFKIDDPVPGGTNKLISIPFVYRIRKVK
VEWPCSPITQDGRGVSTAVILDNFTKATLYDPIYVNSRHTIPQFSTHSRY
FTPKVLDSITDIFQPNKRNQLWRLQTSNDVHVLGAAAFENSKYDQDYNIRVTY
VQREFNLKDPPLKP"
BASE COUNT 451 a 361 c 495 g 461 t
ORIGIN

alignment_scores
Quality: 314.00 Length: 59
Ratio: 5.322 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.305

alignment_block
US-09-209-961-20 x AF055392 ..
Align seg 1/1 to: AF055392 from: 1 to: 1768
1 MetTyrThrSerLeuTrpGlyHisLeuGlyValValLysAlaAsnGlyLe 17
|||||
900 ATGTACACGTCATCTTGGGGCCACCCTGGGTGTGTAAGCAAAATGGGCT 949
17 uLeuTleLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrs 34
|||||
950 GCTAATTTCAGACCGGAAACCCATACCTAGTAAACCACTAGAACAA 999
34 exGlyGlyMetValThrMetValLysLysTrpLeuLeuMetThrPhe 50
|||||
1000 GGTGGTGGATGCTTACCATGGTGAGAGAGTGTGTATGATGACTTTT 1049
51 MetAlaGlyCysArgGlyMetIleTyr 59
1050 ATGGCTGGCTGCCGTGGATGATCTAC 1076

seq_name: gb_vi:AF109397
seq_documentation_block:
LOCUS AF109397 1768 bp DNA circular VRL 02-FEB-1999
DEFINITION Bovine circovirus, complete genome.
ACCESSION AF109397
NID 94210313
VERSION AF109397.1 GI:4210313
KEYWORDS
SOURCE bovine circovirus.
ORGANISM bovine circovirus
REFERENCE 1 (bases 1 to 1768)
Hamel,A.L. and Nayar,G.P.S.
Nucleotide sequence of a circovirus detected in cattle with various
clinical syndromes
Unpublished
2 (bases 1 to 1768)
Hamel,A.L. and Nayar,G.P.S.
Direct Submission
Submitted (27-NOV-1998) Virology Laboratory, Veterinary Diagnostic
Laboratory, Manitoba Agriculture, 545 University Crescent,
Winnipeg, Manitoba R3T 5S6, Canada
3 (bases 1 to 1768)
Hamel,A.L. and Nayar,G.P.S.
Direct Submission
Submitted (02-FEB-1999) Virology Laboratory, Veterinary Diagnostic
Laboratory, Manitoba Agriculture, 545 University Crescent,
Winnipeg, Manitoba R3T 5S6, Canada
Sequence update by submitter
On Feb 2, 1999 this sequence version replaced gi:4028609.
Location/Qualifiers
1..1768
/organism="Bovine circovirus"
/virus
/db_xref="taxon:85542"
/note="BCV"
sequence was obtained from several overlapping PCRs using
DNA extracted from various tissues from cattle; similar to
Porcine circovirus sequence presented in GenBank Accession
Number AF027217"
repeat_region
13..35
/rpt_type=tandem
/rpt_unit=13..18
51..995
/note="ORF-1; similar to Porcine circovirus ORF1 encoded
by the sequence presented in GenBank Accession Number
AF027217."
/codon_start=1
/product="putative Rep and coat protein"
/protein_id="AAD11928.1"
/db_xref="PID:94210314"
/db_xref="GI:4210314"
/translation="MPSKNGSGPQPHRRVFTLNNPSEDEKKIRELPISLFDYFI
VGEENEGRTPHLOGFANFVKQTFNKVYLGARCHEKAKGTQONKEYCSKEN
LIEGAPESQGRSDLSAVSTLLESGLTVAEQHPVTFVFNELGLAELKVGKM
QKRDKNVHVIVPGCGCKSKWAANFADPETTYWPKPNKWDGYPHGEVWIDFY
GWLWDDLRLCDRIYPLTVETGGVTFVFLARSILITSNQTPLEWISSTAVPAVEALYR
RITSLVFNKKNATEQSTEGGQFVTLSPPCPEFPYEINY"
complement(357..671)
/note="similar to Porcine circovirus ORF-3 encoded by the
sequence presented in GenBank Accession Number AF027217."
/codon_start=1
/product="ORF-3"
/protein_id="AAD11930.1"
/db_xref="PID:94210315"
/db_xref="GI:4210315"
/translation="MTVTPVLSRWFVCGFRVCKISSPFAFTTTPRPHNDVYIGLPI
TLHHPFAHFQKFSQPAETSDKRYRVLNCGHOTPALQOQTHSSRQVTPLSLRSRSTF
NK"
complement(386..565)
/note="similar to Porcine circovirus ORF-4 encoded by the
sequence presented in GenBank Accession Number AF027217."
/codon_start=1
/product="ORF-4"
/protein_id="AAD11931.1"
/db_xref="PID:94210316"
/db_xref="GI:4210316"
/translation="MTCFLVFSRFCIFPLTFKSSASPRKFLTNVTCGCSATVRLPL
SNKVLTAVDRSLRCP"
553..732
/codon_start=1
/product="ORF-12"
/protein_id="AAD11927.1"
/db_xref="PID:94028610"

```

```

/db_xref="GI:4028610"
/translation="MTSLWGLGVVAVKANGLLILQRPHTGNHLETSGMVTWVKK
LLMTFMAGCRGMIV"
complement(588..753)
/codon_start=1
/product="ORF-8"
/protein_id="AAD11935.1"
/db_xref="PID:g4028618"
/db_xref="GI:4028618"
/translation="MDIDHTSVVDHETAASHKSHQ"
933..988
complement(989..1033)
/codon_start=1
/product="ORF-11"
/protein_id="AAD11937.1"
/db_xref="PID:g4028620"
/db_xref="GI:4028620"
/translation="MNNKNHYEVIKTKQ"
1016..1177
/codon_start=1
/product="ORF-5"
/protein_id="AAD11932.1"
/db_xref="PID:g4028615"
/translation="MVFIIHLGFKWGVFKIKFSELYINGYTDIVLWVFTVFSR
AEEA"
YVHISRGU"
complement(1022..1027)
complement(1256..1735)
/notes="similar to Porcine circovirus ORF-2 encoded by the
sequence presented in GenBank Accession Number AF027217."
/codon_start=1
/product="ORF-2"
/protein_id="AAD11929.1"
/db_xref="PID:g4028612"
/db_xref="GI:4028612"
/translation="MTYPRRRYRRRTRPRSHGQILRRPWLPHRPHRYRWRKNGI
FNTLRRTGYTVAITVTPSVAVDMRFNIDDFYPPGGTNKISIPFYRIRVKV
VEFWPCSPITQDGRGVSTAVILDDNFKATALTYDPYVYSSRTIIPQFYSYHR"
complement(1522..1611)
/codon_start=1
/product="ORF-6"
/protein_id="AAD11933.1"
/db_xref="PID:g4028616"
/db_xref="GI:4028616"
/translation="MASTPASPAPSDILSRLPQSERPPGRWT"
1524..1631
/codon_start=1
/product="ORF-10"
/protein_id="AAD11936.1"
/db_xref="PID:g4028619"
/db_xref="GI:4028619"
/translation="MSTAEGVLTVALVYPKVRRLVKMPFFLLOR"
1682..1741
/codon_start=1
/product="ORF-7"
/protein_id="AAD11934.1"
/db_xref="PID:g4028617"
/db_xref="GI:4028617"
/translation="MAAGAGSSAVTPPWIRHS"
join(1750..1768,1..13)
join(1762..1768,1..2)
/notes="putative; similar to the nonanucleotide motif of
porcine circovirus presented in GenBank Accession Numbers
AF027217 and U49186"
452 a 362 c 493 g 461 t
BASE COUNT
ORIGIN

```

alignment\_scores:  
 Quality: 314.00 Length: 59  
 Ratio: 5.322 Caps: 0  
 Percent Similarity: 100.000 Percent Identity: 98.305

```

alignment_block:
US-09-209-961-20 x AF109397
Align seg 1/1 to: AF109397 from: 1 to: 1768

1 MetTyrThrSerLeuTrpGlyHisLeuGlyValValLysAlaAsnGlyLe 17
553 ATGTACACGTCATTGTGGGGCCACCTGGTGTGGTAAAGCAATGGGCT 602
17 uleuileLeuGlnThrArgLysProHisThrClyAsnHisLeuLysThr 34
603 GCTAATTTTGCAGACCCGGAACACACATCTGGAACACCACTAGAACAA 652
34 erGlyGlyMetValThrMetValLysLysTrpLeuLeuMetThrPhe 50
653 GTGGTGGGATGGTTACCATGGTGAAGAAGTGGTGTATTGATGACTTTT 702
51 MetAlaGlyCysArgGlyMetIleTyr 59
703 ATGGCTGGCTGCCGTGGGATGATCTAC 729

seq_name: gb_vi:AF109399

seq_documentation_block:
LOCUS AF109399 1768 bp DNA circular VRL 06-JAN-1999
DEFINITION Porcine circovirus type 2-E, complete genome.
ACCESSION AF109399
NID 94108896
VERSION AF109399.1 GI:4106896
KEYWORDS
SOURCE
ORGANISM
porcine circovirus type 2-E.
porcine circovirus type 2-E
Viruses; ssDNA viruses; Circoviridae; Circovirus; #69#line
circovirus type 2
REFERENCE
1 (bases 1 to 1768)
Hamel,A.L. and Nayar,G.P.S.
Nucleotide sequence of four different isolates of circovirus
detected in pigs with various clinical syndromes
Unpublished
REFERENCE
2 (bases 1 to 1768)
Hamel,A.L. and Nayar,G.P.S.
Direct Submission
Submitted (27-NOV-1998) Virology Laboratory, Veterinary Diagnostic
Laboratory, Manitoba Agriculture, 545 University Crescent,
Winnipeg, Manitoba R3T 5S6, Canada
FEATURES
source
1..1768
/organism="porcine circovirus type 2-E"
/virion
/db_xref="taxon:85544"
/notes="similar to Porcine circovirus sequence presented in
GenBank Accession Number AF027217; type-E designation is
based upon restriction endonuclease digestion pattern;
sequence obtained from several overlapping PCRs using DNA
extracted from lung, mesenteric lymph node and tonsil of
pig"
repeat_region 13..35
/rpt_type=tandem
/rpt_unit=13..18
51..995
CDS
1..995
/contig="ORF-1"
/codon_start=1
/product="putative Rep and coat protein"
/protein_id="AAD03071.1"
/db_xref="PID:g4106897"
/db_xref="GI:4106897"
/translation="MPSKKNRSGPQPHKRWFTLNPSDEKRIKELPISLFDYFI
VGEENEGRTPHLQGFANFKVKKQTFNKVYLGARCHIEKAKTDOQNKYCSKEGN
LLIECAPRSQGRSDLSAVTSLLESGSLVTVAEQHPVTFVRNFRGLAEKLKVSGRK
QKRDKNTNHHVIVGPPCGSKSWAANFADPTTYKPKPRKRWGDIHGEVYVDDFY
GWLFWDDLLRLCDRIPLTVETKGTGTFVFLARSILITSNQTPLEWTPSTAVPAVALYR
RITSLVFWKNATEQSTEEGGQFVTLSPCPPEFFPEYNY"
327..332
polya_signal

```

```

CDS
complement(357..671)
/codon_start=1
/product="ORF-3"
/protein_id="AAD03073.1"
/db_xref="PID:g4106899"
/db_xref="GI:4106899"
/translation="MTIPPLVSRWFVPCGRVCKISSPFAFTTPRPHNDVYIGLPI
TLHFAHFKFSQPAEISDKRVLLCNGHOTPALQGGTHSSRQVTPLSLRSSSTL
NK"
complement(386..565)
/codon_start=1
/product="ORF-4"
/protein_id="AAD03074.1"
/db_xref="PID:g4106900"
/db_xref="GI:4106900"
/translation="MTCTLVQSRFCIFPLTFKSSAPRFLTNVTCSSATVTRLPL
SNKVLTAVDRLKCP"
553..732
/codon_start=1
/product="ORF-12"
/protein_id="AAD03081.1"
/db_xref="PID:g4106907"
/db_xref="GI:4106907"
/translation="MTSLGHLGVVKANGLLILQTRKPHGTGNHLETSGMWTVMKVK
LILMTFMAGCRGMV"
complement(688..753)
/codon_start=1
/product="ORF-8"
/protein_id="AAD03078.1"
/db_xref="PID:g4106904"
/db_xref="GI:4106904"
/translation="MDIDHTVSDVHDPTAASHKSHQ"
983..988
complement(989..1033)
/codon_start=1
/product="ORF-11"
/protein_id="AAD03080.1"
/db_xref="PID:g4106906"
/db_xref="GI:4106906"
/translation="MNNKNHYEVIKTKQ"
1016..1363
/notes="similar to Porcine circovirus ORF-5 encoded by the
sequence presented in GenBank Accession Number AF027217."
/codon_start=1
/product="ORF-5"
/protein_id="AAD03075.1"
/db_xref="PID:g4106901"
/db_xref="GI:4106901"
/translation="MVFIHLGFKWGVKIKFSELYIHGYTDIVLVVTFVFSRSEA
YVYHISTGLESHPQLIPFVILWLEVINSGIANRGCEVPGVYVGEGLGDCMAGGVYVGV
IGLGCGLYKXV"
complement(1022..1027)
complement(1034..1735)
/codon_start=1
/product="ORF-2"
/protein_id="AAD03072.1"
/db_xref="PID:g4106898"
/db_xref="GI:4106898"
/translation="MTVPRRRRRRRRHPRLGLOILRRRPLVLPBRRHYRWKRNKI
ENARLSFTGYTKATVSPKAVDMRLNLDQFVPPGGTKNKISIPFEYIRYRKI
VEFWPCSPITQDRGVSSAILIDDNFVIRAKTAQTDYPIVNSRHSRHPQPSIHSYK
FTPQVLDSTIDYFQNNKRNQLMRLQTSRNVDRHVLGTAFENSKYDODYNIRVTM
YQRFENLKDPPLPK"
1524..1631
/codon_start=1
/product="ORF-10"
/protein_id="AAD03079.1"
/db_xref="PID:g4106905"
/db_xref="GI:4106905"
/translation="MSTAEGLVTVALTVYPKVRRERALKMPFFLQR"
complement(1528..1611)
/codon_start=1
/product="ORF-6"

CDS
complement(357..671)
/codon_start=1
/product="ORF-3"
/protein_id="AAD03073.1"
/db_xref="PID:g4106899"
/db_xref="GI:4106899"
/translation="MTIPPLVSRWFVPCGRVCKISSPFAFTTPRPHNDVYIGLPI
TLHFAHFKFSQPAEISDKRVLLCNGHOTPALQGGTHSSRQVTPLSLRSSSTL
NK"
complement(386..565)
/codon_start=1
/product="ORF-4"
/protein_id="AAD03074.1"
/db_xref="PID:g4106900"
/db_xref="GI:4106900"
/translation="MTCTLVQSRFCIFPLTFKSSAPRFLTNVTCSSATVTRLPL
SNKVLTAVDRLKCP"
553..732
/codon_start=1
/product="ORF-12"
/protein_id="AAD03081.1"
/db_xref="PID:g4106907"
/db_xref="GI:4106907"
/translation="MTSLGHLGVVKANGLLILQTRKPHGTGNHLETSGMWTVMKVK
LILMTFMAGCRGMV"
complement(688..753)
/codon_start=1
/product="ORF-8"
/protein_id="AAD03078.1"
/db_xref="PID:g4106904"
/db_xref="GI:4106904"
/translation="MDIDHTVSDVHDPTAASHKSHQ"
983..988
complement(989..1033)
/codon_start=1
/product="ORF-11"
/protein_id="AAD03080.1"
/db_xref="PID:g4106906"
/db_xref="GI:4106906"
/translation="MNNKNHYEVIKTKQ"
1016..1363
/notes="similar to Porcine circovirus ORF-5 encoded by the
sequence presented in GenBank Accession Number AF027217."
/codon_start=1
/product="ORF-5"
/protein_id="AAD03075.1"
/db_xref="PID:g4106901"
/db_xref="GI:4106901"
/translation="MVFIHLGFKWGVKIKFSELYIHGYTDIVLVVTFVFSRSEA
YVYHISTGLESHPQLIPFVILWLEVINSGIANRGCEVPGVYVGEGLGDCMAGGVYVGV
IGLGCGLYKXV"
complement(1022..1027)
complement(1034..1735)
/codon_start=1
/product="ORF-2"
/protein_id="AAD03072.1"
/db_xref="PID:g4106898"
/db_xref="GI:4106898"
/translation="MTVPRRRRRRRRHPRLGLOILRRRPLVLPBRRHYRWKRNKI
ENARLSFTGYTKATVSPKAVDMRLNLDQFVPPGGTKNKISIPFEYIRYRKI
VEFWPCSPITQDRGVSSAILIDDNFVIRAKTAQTDYPIVNSRHSRHPQPSIHSYK
FTPQVLDSTIDYFQNNKRNQLMRLQTSRNVDRHVLGTAFENSKYDODYNIRVTM
YQRFENLKDPPLPK"
1524..1631
/codon_start=1
/product="ORF-10"
/protein_id="AAD03079.1"
/db_xref="PID:g4106905"
/db_xref="GI:4106905"
/translation="MSTAEGLVTVALTVYPKVRRERALKMPFFLQR"
complement(1528..1611)
/codon_start=1
/product="ORF-6"

polyA_signal
CDS
1 MetTyrThrSerLeuTrpGlyHisIeuGlyValValIysAlaAsnGlyLe 17
|||||
553 ATGTACACGTCATGTGGGGCCACCTGGGTGTGTTAAAGCAATGGGCT 602
17 uLeuIleLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrS 34
|||||
603 GCTAATTTTCAGACCCGGAACACACATACCTGGAACACCTAGAACAA 552
34 eGlyGlyMetValThrMetValLysLysTrpLeuLeuLeuMetThrPhe 50
|||||
653 GTGGTGGGATGTTACATGTTGGAAGAGTGGTGTGTTATTGATGACTTTT 702
51 MetAlaGlyCysArgGlyMetIleTyr 59
|||||
703 ATGGCTGGCTGCCGTGGGATGATCTAC 729
seq_name: gb_vi:AF147751

seq_documentation_block:
LOCUS AF147751 1768 bp DNA 08-JUN-1999
DEFINITION Porcine circovirus type 2 unknown genes.
ACCESSION AF147751
NID G5007010
VERSION AF147751.1 GI:5007010
KEYWORDS
SOURCE
ORGANISM
porcine circovirus type 2.
porcine circovirus type 2
Viruses; ssDNA viruses; Circoviridae; Circovirus.
REFERENCE
1 (bases 1 to 1768)
Pogranichnyy, R., Yoon, K.-J., Harms, P., Swenson, S., Zimmerman, J. and
Sorden, S.
Characterization of clinical and immune responses in young swine to
experimental porcine circovirus type II infection
Unpublished
2 (bases 1 to 1768)
Yoon, K.-J. and Pogranichnyy, R.M.
Direct Submission
Submitted (30-APR-1999) Veterinary Diagnostic Laboratory, Iowa
State University, College of Veterinary Medicine, Ames, Iowa 50011,
USA
FEATURES
Location/Qualifiers
1..1768
/organism="porcine circovirus type 2"

BASE COUNT 448 a 362 c 495 g 463 t
ORIGIN

alignment_scores:
Quality: 314.00 Length: 59
Ratio: 5.322 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.305

alignment_block:
US-09-209-961-20 x AF109399 ..
Align seg 1/1 to: AF109399 from: 1 to: 1768
1 MetTyrThrSerLeuTrpGlyHisIeuGlyValValIysAlaAsnGlyLe 17
|||||
553 ATGTACACGTCATGTGGGGCCACCTGGGTGTGTTAAAGCAATGGGCT 602
17 uLeuIleLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrS 34
|||||
603 GCTAATTTTCAGACCCGGAACACACATACCTGGAACACCTAGAACAA 552
34 eGlyGlyMetValThrMetValLysLysTrpLeuLeuLeuMetThrPhe 50
|||||
653 GTGGTGGGATGTTACATGTTGGAAGAGTGGTGTGTTATTGATGACTTTT 702
51 MetAlaGlyCysArgGlyMetIleTyr 59
|||||
703 ATGGCTGGCTGCCGTGGGATGATCTAC 729
seq_name: gb_vi:AF147751

seq_documentation_block:
LOCUS AF147751 1768 bp DNA 08-JUN-1999
DEFINITION Porcine circovirus type 2 unknown genes.
ACCESSION AF147751
NID G5007010
VERSION AF147751.1 GI:5007010
KEYWORDS
SOURCE
ORGANISM
porcine circovirus type 2.
porcine circovirus type 2
Viruses; ssDNA viruses; Circoviridae; Circovirus.
REFERENCE
1 (bases 1 to 1768)
Pogranichnyy, R., Yoon, K.-J., Harms, P., Swenson, S., Zimmerman, J. and
Sorden, S.
Characterization of clinical and immune responses in young swine to
experimental porcine circovirus type II infection
Unpublished
2 (bases 1 to 1768)
Yoon, K.-J. and Pogranichnyy, R.M.
Direct Submission
Submitted (30-APR-1999) Veterinary Diagnostic Laboratory, Iowa
State University, College of Veterinary Medicine, Ames, Iowa 50011,
USA
FEATURES
Location/Qualifiers
1..1768
/organism="porcine circovirus type 2"

```

/isolate="ISUVDL 98-15237"  
/specific\_host="Sus scrofa"  
/db\_xref="taxon:85708"  
/note="Isolated from pigs with postweaning multisystemic  
wasting syndrome"

CDS

51..995  
/note="ORF1; 35-36kd"  
/codon\_start=1  
/product="unknown"  
/protein\_id="AAD37776.1"  
/db\_xref="PID:G5007011"  
/db\_xref="GI:5007011"  
/translation="MPSKNGRSQPQHRKRWFTLNNPSEDERKKIRELPISLFDYFI  
VGEENEGRTPLHGFANFKVOTFNKVKYLGAICHIEKAKGTDOONKEVCSKGN  
LLIEGAPSGORSGLSTAVSLLESGLVVAQHPVTFVNRFGLAELLYKSGKM  
QKRDKNVHVIVGPPGCKSKWANFADPETYWKPPRNKWDGVLHGEVVVIDFY  
GWLPMDDLRLCDRYPLTVETKGTVPFLARSILITSNQTPLEWYSSTAVPAVEALYR  
RITSLVFWKNATEQSTEGGQFVTLSPPCPEFFPEYNY"  
complement(1034..1735)  
/note="ORF2; 27-28kd"  
/codon\_start=1  
/product="unknown"  
/protein\_id="AAD37777.1"  
/db\_xref="PID:G5007012"  
/db\_xref="GI:5007012"  
/translation="MTYPRRYRRRHRPRSHLGOILRRPWLHPHRYRWRKNGI  
FNRLSRTPGYVKAITVTPSNVDMMRFNIDDFVPPGGGTNKISIPREYVIRVK  
VWFPCSPITQDGRGVSTAVILDDNFNTALTYDIPVNYSSRTIIPQPSCHRY  
FPPKPLVDSTIDYFQPNKRNQLWLRLOTSRNVHDVGLGTAFENSIIYDQDNIIRVTMY  
VQFREFNLKDPPLKP"

CDS

450 a 362 c 494 g 462 t  
BASE COUNT  
ORIGIN

alignment\_scores:  
Quality: 314.00 Length: 59  
Ratio: 5.322 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 98.305  
alignment\_block:  
US-09-209-961-20 x AF147751 ..  
Align seg 1/1 to: AF147751 from: 1 to: 1768

1 MetTyrThrSerLeuTrpGlyHisLeuGlyValValLysAlaAsnGlyLe 17  
|||||  
553 ATGTACACGTCATTGTGGGGCCACCTGGGTGTGTTAAAGCAAAATGGGCT 602  
17 uLeuileLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrS 34  
|||||  
603 GCTAATTTTGCAGACCCGGAACCACTACTGGAACCAACCACTAGAACAA 652  
34 erGlyGlyMetValThrMetValLysLysTrpLeuLeuLeuMetThrPhe 50  
|||||  
653 GTGGTGGGATGTTACCATGTTGAAGAGTGGTGTGTTATTGATGACTTTT 702  
51 MetAlaGlyCysArgGlyMetIleTyr 59  
|||||  
703 ATGGCTGGCTGCCGTGGGATGATCTAC 729

seq\_name: gb\_vi:PCAJ3185

seq\_documentation\_block:  
LOCUS PCAJ3185 1768 bp DNA circular VRL 03-JUL-1998  
DEFINITION Porcine circovirus DNA, complete genome, isolate ISU-31.  
ACCESSION AJ223185  
NID 93293040  
VERSION AJ223185.1 GI:3293040  
KEYWORDS complete genome.  
SOURCE porcine circovirus.  
ORGANISM porcine circovirus.  
REFERENCE 1 (bases 1 to 1768)

AUTHORS  
TITLE  
JOURNAL

Morozov, I.  
Direct Submission  
Submitted (02-JAN-1998) Morozov I., Veterinary Medical Research  
Institute, Iowa State University, 1802 Elwood Dr., VMRI Ames, Iowa  
50011, U.S.A

REFERENCE

2 (bases 1 to 1768)  
Morozov, I., Sirinarumitr, T., Sorden, S.D., Halbur, P.G., Morgan, M.K.,  
Yoon, K.J. and Paul, P.S.

AUTHORS

Detection of a novel strain of Porcine circovirus in pigs with  
postweaning multisystemic Wasting Syndrome (PMWS)

TITLE

Unpublished

JOURNAL

Location/Qualifiers

FEATURES

source  
1..1768  
/organism="porcine circovirus"  
/virion  
/isolate="ISU-31"  
/specific\_host="Sus scrofa"  
/db\_xref="taxon:46221"  
51..995  
/note="ORF1"  
/codon\_start=1  
/protein\_id="CAAL1157.1"  
/db\_xref="PID:el310034"  
/db\_xref="PID:g3293041"  
/db\_xref="GI:3293041"  
/translation="MPSKNGRSQPQHRKRWFTLNNPSEDERKKIRELPISLFDYFI  
VGEENEGRTPLHGFANFKVOTFNKVKYLGAICHIEKAKGTDOONKEVCSKGN  
LLIEGAPSGORSGLSTAVSLLESGLVVAQHPVTFVNRFGLAELLYKSGKM  
QKRDKNVHVIVGPPGCKSKWANFADPETYWKPPRNKWDGVLHGEVVVIDFY  
GWLPMDDLRLCDRYPLTVETKGTVPFLARSILITSNQTPLEWYSSTAVPAVEALYR  
RITSLVFWKNATEQSTEGGQFVTLSPPCPEFFPEYNY"  
51..995  
/note="ORF1"  
/codon\_start=1  
/protein\_id="CAAL1157.1"  
/db\_xref="PID:el310034"  
/db\_xref="PID:g3293041"  
/db\_xref="GI:3293041"  
/translation="MPSKNGRSQPQHRKRWFTLNNPSEDERKKIRELPISLFDYFI  
VGEENEGRTPLHGFANFKVOTFNKVKYLGAICHIEKAKGTDOONKEVCSKGN  
LLIEGAPSGORSGLSTAVSLLESGLVVAQHPVTFVNRFGLAELLYKSGKM  
QKRDKNVHVIVGPPGCKSKWANFADPETYWKPPRNKWDGVLHGEVVVIDFY  
GWLPMDDLRLCDRYPLTVETKGTVPFLARSILITSNQTPLEWYSSTAVPAVEALYR  
RITSLVFWKNATEQSTEGGQFVTLSPPCPEFFPEYNY"

CDS

BASE COUNT 452 a 361 c 492 g 463 t  
ORIGIN

alignment\_scores:  
Quality: 314.00 Length: 59  
Ratio: 5.322 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 98.305  
alignment\_block:  
US-09-209-961-20 x PCAJ3185 ..  
Align seg 1/1 to: PCAJ3185 from: 1 to: 1768

1 MetTyrThrSerLeuTrpGlyHisLeuGlyValValLysAlaAsnGlyLe 17  
|||||  
553 ATGTACACGTCATTGTGGGGCCACCTGGGTGTGTTAAAGCAAAATGGGCT 602  
17 uLeuileLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrS 34  
|||||  
603 GCTAATTTTGCAGACCCGGAACCACTACTGGAACCAACCACTAGAACAA 652  
34 erGlyGlyMetValThrMetValLysLysTrpLeuLeuLeuMetThrPhe 50  
|||||  
653 GTGGTGGGATGTTACCATGTTGAAGAGTGGTGTGTTATTGATGACTTTT 702  
51 MetAlaGlyCysArgGlyMetIleTyr 59  
|||||  
703 ATGGCTGGCTGCCGTGGGATGATCTAC 729

seq\_name: gb\_vi:AF117753

seq\_documentation\_block:  
LOCUS AF117753 1768 bp DNA circular VRL 04-FEB-1999  
DEFINITION Porcine circovirus type 2-D, complete genome.  
ACCESSION AF117753  
NID 94219093  
VERSION AF117753.1 GI:4219093  
KEYWORDS porcine circovirus type 2-D.  
SOURCE porcine circovirus type 2-D.  
ORGANISM porcine circovirus type 2-D.  
REFERENCE 1 (bases 1 to 1768)

```

circovirus type 2.
1 (bases 1 to 1768)
AUTHORS Hamel,A.I. and Nayar,G.P.S.
TITLE Genetic characterization of four novel type-2 Porcine circoviruses
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1768)
AUTHORS Hamel,A.I. and Nayar,G.P.S.
TITLE Direct Submission
JOURNAL Submitted (05-JAN-1999) Virology Laboratory, Manitoba Agriculture,
545 University Crescent, Winnipeg, Manitoba R3T 5S6, Canada
FEATURES
    source
        Location/Qualifiers
            1..1768
                /organism="porcine circovirus type 2-D"
                /virion
                /db_xref="taxon:86385"
                /note="type 2-D designation is based upon restriction
                endonuclease digestion pattern
                several overlapping PCR fragments were sequenced; virus
                isolated from lung, mesenteric lymph node and tonsil
                tissue; similar to Porcine circovirus sequence presented
                in GenBank Accession Number AF027217"
repeat_region
    13..35
        /rpt_type=tandem
        /rpt_unit=13..18
        51..995
            /note="putative Rep and coat protein"
            /codon_start=1
            /product="ORF-1"
            /protein_id="AAD12308.1"
            /db_xref="PID:g4219094"
            /db_xref="GI:4219094"
            /translation="MPSKNGRSGPOPKHWFTLNNPSEDEKKIRELPISLFDYFI
            VGEENEGRTPHLOGFANFYKQTKENKYNWLGACSHIEKAGTQONKEYSKEN
            LMECAPRSQGRSDLSAVSLTSSGSLTVVAEQHPVTFVFNFRGLAELLKVSNGM
            QKRDKNVHVPPGCGSKWAANFADPTTYWKPNNKWDGHEVVYIDDFY
            GWPDLRLCLDRYPLTETKGTGTFVFLARSILITSNQTPLEWYSSAAVPAVEALYR
            RITSLVFNKATEQSTEGGQFVTLSPPCPEFFEYNY"
            327..332
                complement(357..671)
                /codon_start=1
                /product="ORF-3"
                /protein_id="AAD12309.1"
                /db_xref="PID:g4219095"
                /db_xref="GI:4219095"
                /translation="MTYIPLSRNFPVCGFRVCKISSPPAFATPRWPHNDVIRLPI
                TLLHFAHFQKFSQPAEISDKRYRVLCCNGHOTPAKQGTHSGRQVTPLSLRSSSTF
                HQ"
                complement(386..565)
                /codon_start=1
                /product="ORF-4"
                /protein_id="AAD12310.1"
                /db_xref="PID:g4219096"
                /db_xref="GI:4219096"
                /translation="MTCTFVQSFRFCIFPLTFKSSASPRKFLNVTGCCSATVTRLPL
                SSKVLTAVDRLRCP"
                553..732
                    /codon_start=1
                    /product="ORF-12"
                    /protein_id="AAD12311.1"
                    /db_xref="PID:g4219097"
                    /db_xref="GI:4219097"
                    /translation="MTSLWHLGVAKANGLLILQTRKPHGTGNHLETSGGMVTWKKW
                    LLLMTFMAGCRGMIV"
                    complement(688..753)
                    /codon_start=1
                    /product="ORF-8"
                    /protein_id="AAD12312.1"
                    /db_xref="PID:g4219098"
                    /db_xref="GI:4219098"
                    /translation="MDTDHTVSVDPHTAASHKSHQ"
                    983..988
                        complement(989..1033)
                        /codon_start=1
polyA_signal
CDS
    /product="ORF-11"
    /protein_id="AAD12313.1"
    /db_xref="PID:g4219099"
    /db_xref="GI:4219099"
    /translation="MNNKNHYEVIKKTQ"
    1016..1177
        /codon_start=1
        /product="ORF-5"
        /protein_id="AAD12314.1"
        /db_xref="PID:g4219100"
        /db_xref="GI:4219100"
        /translation="MVFIHLGKVGKIKFSELYIHGYTDIVVLVYVTVFERSAEA
        YMYVSSSL"
        complement(1022..1027)
        complement(1256..1735)
        /note="similar to Porcine circovirus ORF-2 product encoded
        by the sequence presented in GenBank Accession Number
        AF027217"
        /codon_start=1
        /product="ORF-2"
        /protein_id="AAD12315.1"
        /db_xref="PID:g4219101"
        /db_xref="GI:4219101"
        /translation="MTFPFRRYRRRRHRRPSHLGQILRRRPWFLEHPRHRYRRWRKNGI
        ENRLSRTFGYTVKATVTPSWAVDMLEFKIDDFVPPGGTGNKISIPFEYVIRKVK
        VEFWPCSPITQDGRGVGSSAVILDNFFPKSTALTYDPVYVYSSRHTITQPFYHSR"
        1524..1631
            /codon_start=1
            /product="ORF-10"
            /protein_id="AAD12317.1"
            /db_xref="PID:g4219103"
            /db_xref="GI:4219103"
            /translation="MSTAQEGVLTVVTPVKVRERVLKMPFELLOR"
            complement(1528..1611)
            /product="ORF-6"
            /protein_id="AAD12316.1"
            /db_xref="PID:g4219102"
            /db_xref="GI:4219102"
            /translation="MASSTPASPAPSDILSLRQPSPGPR"
            1682..1741
                /codon_start=1
                /product="ORF-7"
                /protein_id="AAD12318.1"
                /db_xref="PID:g4219104"
                /db_xref="GI:4219104"
                /translation="MAAGAVSSSAVTPPWKRHS"
                join(1750..1768,1..13)
                /note="putative replication site"
                join(1762..1768,1..2)
                /note="similar to the nonanucleotide motif of Porcine
                circovirus presented in GenBank Accession Numbers AF027217
                and U49186"
                452 a 358 c 497 g 461 t
                BASE COUNT
                ORIGIN
alignment_scores:
    Quality: 310.00 Length: 59
    Ratio: 5.254 Gaps: 0
    Percent Similarity: 100.000 Percent Identity: 96.610
alignment_block:
    US-09-209-961-20 x AF117753 ..
Align seg 1/1 to: AF117753 from: 1 to: 1768
1 MetYrThrSerLeuTrpGlyHisLeuGlyValValysAlaAsnGlyLe 17
|||||
553 ATGTACACGTCATTGTGGGGCCACCTGGGTGGTGGCAAAAGCAATGGCT 602
|||||
17 uleuileLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrs 34
|||||

```

603 GCTAATTTTCAGACCCGGAACACACATCTGGAACACCACTAGAACAA 652  
 34 erGlyGlyMetValThrMetValLysLysTrpLeuLeuMetThrPhe 50  
 653 GTGGTGGGATGGTTACCATGGTGAAGAAGTGGTGTATTGATGACTTTT 702

51 MetAlaGlyCysArgGlyMetIleTyr 59  
 703 ATGGCTGGCTGCCGTGGGATGATCTAC 729

seq\_name: gb\_vi:AF154679

seq\_documentation\_block: 1768 bp DNA circular VRL 13-JUN-1999  
 LOCUS AF154679  
 DEFINITION Porcine circovirus, complete genome.  
 ACCESSION AF154679  
 NID 95052004  
 VERSION AF154679.1 GI:5052004  
 SOURCE  
 ORGANISM  
 porcine circovirus.  
 Viruses: ssDNA viruses; Circoviridae; Circovirus.  
 REFERENCE  
 1 (bases 1 to 1768)  
 AUTHORS Kuo,T.Y., Chou,Y.C. and Lai,S.S.  
 TITLE Complete nucleotide sequences analysis of porcine circovirus outbreak in Taiwan  
 JOURNAL Unpublished  
 REFERENCE  
 2 (bases 1 to 1768)  
 AUTHORS Kuo,T.Y., Chou,Y.C. and Lai,S.S.  
 TITLE Direct Submission  
 JOURNAL Submitted (27-MAY-1999) Veterinary Medicine, National Taiwan University, 142, Chousan Road, Taipei 106, Taiwan

FEATURES  
 source  
 1..1768  
 /organism="porcine circovirus"  
 /isolate="MLTW98"  
 /db\_xref="taxon:46221"  
 /country="Taiwan"

CDS  
 51..995  
 /note="putative Rep protein"  
 /codon\_start=1  
 /product="ORF1"  
 /protein\_id="AAD38398.1"  
 /db\_xref="PID:g5052005"  
 /db\_xref="GI:5052005"  
 /translation="MPSKNGRSQPQPKRWVFTLNPSDEKKIRLPLSLDFCFI  
 VEEGNEGRTPHLOGFANFKVKKOTFNKVFYFGARHIEKAKGTDQNKKEYCSKGN  
 LLIEGAPRSQGSQSLSTAVSTLSEGLSVTAHQHPVTFVFNFRGLAELLKVSQGM  
 OKRDKNTNHHVIVGPPCGKSKWAANFADPTTYWKPNNKWDYHGHEEVVVDIFY  
 GWLPDLDLLRCDRYPLTVETKGTVPFLARSLITSNQTPLEWYSSITAVPAVEALYR  
 RITSLVFNKNATEQSTEGGQFVTLSPCPPEFYINY"  
 BASE COUNT 451 a 367 c 495 g 455 t  
 ORIGIN

alignment\_scores:  
 Quality: 310.00 Length: 59  
 Ratio: 5.254 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 96.610

alignment\_block:  
 US-09-209-961-20 x AF154679 ..

Align seg 1/1 to: AF154679 from: 1 to: 1768

1 MetTyrThrSerLeuTprGlyHisLeuGlyValLysLysAlaAsnGlyLe 17  
 553 ATGTACACGTCATCTGGGGCCACCTGGGTGGTGAAGCAATGGGCT 602

17 uLeuIleLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrS 34  
 603 GCTAATTTTCAGACCCGGAACACACATCTGGAACACCACTCGAACAA 652

34 erGlyGlyMetValThrMetValLysLysTrpLeuLeuMetThrPhe 50  
 653 GTGGTGGGATGGTTACCATGGTGAAGAAGTGGTGTATTGATGACTTTT 702

51 MetAlaGlyCysArgGlyMetIleTyr 59  
 703 ATGGCTGGCTGCCGTGGGATGATCTAC 729

seq\_name: gb\_vi:AF166528

seq\_documentation\_block: 1768 bp mRNA 19-AUG-1999  
 LOCUS AF166528  
 DEFINITION Porcine circovirus complete genome.  
 ACCESSION AF166528  
 NID 95739338  
 VERSION AF166528.1 GI:5739338  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 porcine circovirus.  
 Viruses: ssDNA viruses; Circoviridae; Circovirus.  
 REFERENCE  
 1 (bases 1 to 1768)  
 AUTHORS Yang,K.H., Lee,Y.F., Chao,D.S., Shieh,Y.C. and Lai,S.S.  
 TITLE Complete nucleotide sequences of porcine circovirus Tainan strand outbreak in Taiwan  
 JOURNAL Unpublished  
 REFERENCE  
 2 (bases 1 to 1768)  
 AUTHORS Yang,K.H., Lee,Y.F., Chao,D.S., Shieh,Y.C. and Lai,S.S.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-JUL-1999) Veterinary Medicine, National Chia-Yi Institute of Technology, 300 Shai Fu Road, Lu Liao Li, Chia-Yi City 600, Taiwan

FEATURES  
 source  
 1..1768  
 /organism="porcine circovirus"  
 /strain="Tainan"  
 /db\_xref="taxon:46221"  
 /country="Taiwan"

CDS  
 51..995  
 /codon\_start=1  
 /product="putative Rep protein"  
 /protein\_id="AAD50432.1"  
 /db\_xref="PID:g5739339"  
 /db\_xref="GI:5739339"  
 /translation="MPSKNGRSQPQPKRWVFTLNPSDEKKIRLPLSLDFYFI  
 VEEGNEGRTPHLOGFANFKVKKOTFNKVFYFGARHIEKAKGTDQNKKEYCSKGN  
 LLIEGAPRSQGSQSLSTAVSTLSEGLSVTAHQHPVTFVFNFRGLAELLKVSQGM  
 OKRDKNTNHHVIVGPPCGKSKWAANFADPTTYWKPNNKWDYHGHEEVVVDIFY  
 GWLPDLDLLRCDRYPLTVETKGTVPFLARSLITSNQTPLEWYSSITAVPAVEALYR  
 RITSLVFNKNATEQSTEGGQFVTLSPCPPEFYINY"  
 BASE COUNT 453 a 367 c 492 g 456 t  
 ORIGIN

alignment\_scores:  
 Quality: 310.00 Length: 59  
 Ratio: 5.254 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 96.610

alignment\_block:  
 US-09-209-961-20 x AF166528 ..

Align seg 1/1 to: AF166528 from: 1 to: 1768

1 MetTyrThrSerLeuTprGlyHisLeuGlyValLysLysAlaAsnGlyLe 17  
 553 ATGTACACGTCATCTGGGGCCACCTGGGTGGTGAAGCAATGGGCT 602

17 uLeuIleLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrS 34  
 603 GCTAATTTTCAGACCCGGAACACACATCTGGAACACCACTCGAACAA 652

34 erGlyGlyMetValThrMetValLysLysTrpLeuLeuMetThrPhe 50

|||||  
653 GTGGTGATGTTACCATGCTGAAGAAGTGGTTGTTATTGACTTTT 702

51 MetalaGlyCysArgGlyMetIleTyr 59  
|||||  
703 ATGTTGGTCCCGGGGATGATCAC 729

seq\_name: gb\_vi:AF055393

seq\_documentation\_block: 1767 bp DNA circular VRL 13-SEP-1998  
LOCUS AF055393 Porcine circovirus Type II from France, complete genome.  
DEFINITION Porcine circovirus Type II from France, complete genome.  
ACCESSION AF055393  
NID q3598820  
VERSION AF055393.1 GI:3598820  
KEYWORDS porcine circovirus.  
SOURCE  
ORGANISM Viruses; ssDNA viruses; Circoviridae; Circovirus.  
REFERENCE 1 (bases 1 to 1767)  
AUTHORS Meenan,B.M., McNeill,F., Todd,D., Kennedy,S., Jewhurst,V.A.,  
Ellis,J.A., Hassard,L.E., Clark,E.G., Haines,D.M. and Allan,G.M.  
TITLE Characterization of novel circovirus DNAs associated with wasting  
syndromes in pigs  
JOURNAL J. Gen. Virol. 79 (Pt 9), 2171-2179 (1998)  
MEDLINE 98418498  
REFERENCE 2 (bases 1 to 1767)  
AUTHORS Meenan,B.M.  
TITLE Direct Submission  
JOURNAL Submitted (26-MAR-1998) Brian M Meenan, Veterinary Science, The  
Queen's University of Belfast, Stoney Road, Belfast, Northern  
Ireland, BT4 3SD, U.K.

FEATURES Location/Qualifiers

1..1767  
/organism="porcine circovirus"  
/specific\_host="Sus scrofa"  
/strain="Porcine circovirus Type II"  
/db\_xref="taxon:46221"  
/clone="Imp.1011 48121 EcoRI No.17"  
/country="France"  
complement(101..190)  
/codon\_start=1  
/product="ORF6"  
/protein\_id="AAC35329.1"  
/db\_xref="PID:g3598830"  
/db\_xref="GI:3598830"  
/translation="MASSTPASPAPSDILSSPQSRPPGRWT"  
103..210  
/codon\_start=1  
/product="ORF10"  
/protein\_id="AAC35327.1"  
/db\_xref="PID:g3598828"  
/db\_xref="GI:3598828"  
/translation="KSTAQEGVLTVVRLTVYKVRERVLKMPFFLLQR"  
261..320  
/codon\_start=1  
/product="ORF7"  
/protein\_id="AAC35328.1"  
/db\_xref="PID:g3598829"  
/db\_xref="GI:3598829"  
/translation="MAAGAVSSPVPWPWIRH"  
complement(311..439)  
/codon\_start=1  
/product="ORF9"  
/protein\_id="AAC35325.1"  
/db\_xref="PID:g3598826"  
/db\_xref="GI:3598826"  
/translation="MGLGSASSILLARHVAEVLPRCCRSALVILTAHFFRFQI"  
398..1342  
/codon\_start=1  
/product="putative Rep protein (ORF1)"  
/protein\_id="AAC35320.1"

/db\_xref="PID:g3598821"  
/db\_xref="GI:3598821"  
/translation="MPSKKNRGSPQPKRWVFTLNPSSEDERKKIRDLPLSLDFYFI  
VGEENEGEPTPHLOGFANFVKQTFNKVWYLGARCHIEKAKGTQDNKYECSKEGN  
LMCEGAPRQSGQSDLSAVSTLSEGSVTVAEQHPVIFVRNFRGLAELLKVSGRKM  
QKRDWKTNVHIVGPPGCGSKWAANFADPETTYKPPRNKWDGYHGEYVVIDFY  
GWLFDLLRLCDRYPLTIVTKGTGTPFLARSLITNQTPLEWYSTAVPAVEALYR  
RITSLVFWKNATEQSTEEGGFVTLSPPCPEFPEINY"  
674..679  
complement(704..1018)  
/note="predicted 11.9 kDa protein"  
/codon\_start=1  
/product="ORF3"  
/protein\_id="AAC35322.1"  
/db\_xref="PID:g3598823"  
/db\_xref="GI:3598823"  
/translation="MVTIPPLVSMFVCGFRVCKIISPFAFTTPRNPHNDVYISLPI  
TLLEFFAHFQKFSQPAEISDKRVLLCNHQTALQOQTHSSRQVTPLSLRSSSTL  
HQ"

polyA\_signal  
CDS

complement(733..912)  
/note="predicted 6.5 kDa protein"  
/codon\_start=1  
/product="ORF4"  
/protein\_id="AAC35323.1"  
/db\_xref="PID:g3598824"  
/db\_xref="GI:3598824"  
/translation="MTCTLVFQSRFCIFPLTFKSSASPRKFLTNVTCGCSATVRLPL  
SNKVLTAVDRLRCP"  
complement(1035..1100)  
/note="predicted 2.3 kDa protein"  
/codon\_start=1  
/product="ORF8"  
/protein\_id="AAC35324.1"  
/db\_xref="PID:g3598825"  
/db\_xref="GI:3598825"  
/translation="MDIDHTVSDVDPRAASHKSHQ"

CDS

polyA\_signal  
CDS

1330..1335  
complement(1336..1380)  
/note="predicted 1.8 kDa protein"  
/codon\_start=1  
/product="ORF11"  
/protein\_id="AAC35326.1"  
/db\_xref="PID:g3598827"  
/db\_xref="GI:3598827"  
/translation="MNNKNHYEVIKKTQ"  
complement(1369..1374)  
complement(join(1380..1767,1..314))  
/note="predicted 27.8 kDa protein"  
/codon\_start=1  
/product="ORF2"  
/protein\_id="AAC35321.1"  
/db\_xref="PID:g3598822"  
/db\_xref="GI:3598822"  
/translation="MTYPRRYRRRRHPRSHLQILRRRPLVHPRHRYVRRENGI  
FNRLSKTFCITVKRTVTPVSAVDMRRNINDFLPPGSGNSRSPVFETIRKVK  
VEFPCSPITQDGRGVSSAVIDLDNFVATATYDYPVNYSSRHITQPFYSHSRY  
FTPKVLDSTIDYIPNNKRNQLRLQTAGNVHDVGLGTAFAENSIDYQEIYRVTMY  
VQFREFNLKDPPLNP"

polyA\_site  
CDS

BASE COUNT 447 a 360 c 502 g 458 t  
ORIGIN

alignment\_scores:  
Quality: 307.00 Length: 59  
Ratio: 5.293 Gaps: 0  
Percent Similarity: 98.305 Percent Identity: 96.610

alignment\_block:  
US-09-209-961-20 x AF055393 ..

Align seg 1/1 to: AF055393 from: 1 to: 1767

1 MetTyrThrSerLeuTrpGlyHisLeuGlyValValLysAlaasnGlyLe 17



```
|||||
900 ATGTACACGTCATTGTGGGCCACCTGGGTGGTAAAGCAAATGGGCT 949
17 uLeuIleLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrS 34
|||||
950 GCTAATTTTGCAGACCCGGAACCCACATACCTGGAAACCACTAGAAACAA 999
34 erGlyGlyMetValThrMetValLysLysTripleLeuLeuMetThrPhe 50
|||||
1000 GTGGTGGGATGGTTACCATGGTGGAAGTGGTGTATTATGATGACTTTT 1049
51 MetAlaGlyCysArgGlyMetIleTyr 59
|||||
1050 ATGGCTGGCTGCCCTGGGATGATCTAC 1076
```

OM of: US-09-209-961-20 to: N\_Geneseq\_36:\* out\_format : pfs  
Date: Dec 28, 1999 1:50 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlp  
-Q/cn2.1/USPTO.spool/US09209961/runat\_22121999\_101043\_17942/app\_query.fasta.1  
-DB=N\_Geneseq\_36 -Qfmt=fastap -SUFFIX=ring -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FCAPOP=6.000 -FCAPEXT=7.000 -FCAPOP=10.000 -FCAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotum62  
-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR\_SCORE=escore  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09209961  
-NCPU=6 -ICPU=3 -NO\_XLPXY -WAIT -THREADS=1

Search information block:  
Query: US-09-209-961-20  
Query length: 59  
Database: N\_Geneseq\_36:\*  
Database sequences: 311585  
Database length: 125096042  
Search time (sec): 140.030000

sequence	Strd	Orig	ZScore	Escore	Len	Documentation
N_Geneseq_36:X35012	+	314.00	806.60	2.5e-37	1768	Genomic DNA sequence of PCV st
N_Geneseq_36:X35213	+	314.00	806.60	2.5e-37	1768	Nucleotide sequence of PCV iso
N_Geneseq_36:X35212	+	314.00	806.60	2.5e-37	1768	Nucleotide sequence of PCV iso
N_Geneseq_36:X35214	+	314.00	806.60	2.5e-37	1768	Nucleotide sequence of PCV iso
N_Geneseq_36:X35211	+	307.00	787.91	2.8e-36	1767	Nucleotide sequence of PCV iso
N_Geneseq_36:X35013	+	307.00	787.91	2.8e-36	1767	Nucleotide sequence of PCV iso
N_Geneseq_36:X41449	-	65.00	159.86	0.2716	315	Human secreted protein 5' EST S
N_Geneseq_36:X06052	+	57.50	137.66	4.56	379	Human brain Expressed Sequence
N_Geneseq_36:X40573	+	54.00	136.94	5.01	168	Human secreted protein 5' EST S
N_Geneseq_36:X089836	+	55.50	133.48	6.86	309	Human death associated protein
N_Geneseq_36:X046896	+	51.00	134.04	8.26	114	JEST-39D10 primer B2137CC. Huma
N_Geneseq_36:T31291	-	63.50	130.83	10.96	3269	Rat poly-immunoglobulin recept
N_Geneseq_36:X87541	-	54.50	130.74	11.09	342	EST clone DN293. New polynucleo
N_Geneseq_36:X75294	-	58.00	128.92	14.01	980	Wheat ubiquitin-conjugating enz
N_Geneseq_36:T25800	+	52.50	127.74	16.28	274	Human gene signature HUMGS08028
N_Geneseq_36:T65664	+	50.00	124.95	23.30	190	Polymorphic repeat sequence clc
N_Geneseq_36:T24240	+	53.00	123.77	27.11	452	Human gene signature HUMGS04028
N_Geneseq_36:X78741	-	51.00	123.06	29.69	292	Staphylococcus aureus contig S
N_Geneseq_36:X39589	-	51.50	123.00	29.91	333	Human secreted protein 5' EST S
N_Geneseq_36:X39586	-	52.00	122.58	31.57	393	EST clone AS216. New polynucleo
N_Geneseq_36:X39586	-	51.50	122.51	31.88	349	Human secreted protein 5' EST S
N_Geneseq_36:X22390	+	47.50	121.14	38.00	145	Human HuN6 fragment for product
N_Geneseq_36:T08171	+	56.50	121.07	38.34	1408	Full length coconut LPAAT clon
N_Geneseq_36:T95297	+	58.00	120.84	40.49	2138	Rat glial cell line-derived ne
N_Geneseq_36:X93911	+	58.00	120.64	40.49	2138	Glial cell-line derived neurot
N_Geneseq_36:X08172	+	45.50	120.51	41.20	93	Primer for MMP19 coding sequence
N_Geneseq_36:X39687	+	53.00	120.33	42.13	625	Gastric cancer associated gene
N_Geneseq_36:X4975	+	58.00	119.51	46.79	2378	Rat glial cell derived neurot
N_Geneseq_36:X054052	-	56.00	118.66	52.19	1557	Rat bone formation-inducing pr
N_Geneseq_36:X062384	+	46.00	118.52	52.48	126	biog gene regulatory region (p
N_Geneseq_36:X003333	+	49.00	118.55	52.95	270	Fragment of post-transfusion, n
N_Geneseq_36:X42651	-	49.00	118.55	52.95	270	PT-NANB virus lambda gtl1 clone
N_Geneseq_36:X097712	-	51.00	118.40	53.96	453	Branched-chain alpha-ketoacid
N_Geneseq_36:X20150	+	49.50	118.32	54.55	313	Probe (51) for microbial genes
N_Geneseq_36:X90753	+	56.00	118.22	55.26	1624	Nucleotide sequence of clone M
N_Geneseq_36:T73870	+	60.50	117.26	62.45	5518	Cotton fibre promoter clone 4
N_Geneseq_36:T73865	+	60.50	117.21	62.90	5547	Cotton fibre promoter clone 4
N_Geneseq_36:X41446	+	46.00	116.91	65.32	148	Human secreted protein 5' EST S
N_Geneseq_36:T26682	-	48.50	116.64	67.65	285	Human gene signature HUMGS08932
N_Geneseq_36:T18795	-	56.50	116.21	71.51	2226	Mouse immunophilin clone 213-1
N_Geneseq_36:T69548	-	53.50	116.13	72.19	1053	Rat pheromone receptor VN5 cDN

N\_Geneseq\_36:N90105 - 53.00 116.12 72.35 930 ! Human non-specific cross-rea  
N\_Geneseq\_36:Q21401 + 47.50 115.75 75.85 241 ! Human leukocyte antigen DRw1  
N\_Geneseq\_36:V43035 + 53.00 115.72 76.08 965 ! Streptococcus pneumoniae pol  
  
seq\_name: N\_Geneseq\_36:X35012  
  
seq\_documentation\_block:  
ID X35012 standard; DNA; 1768 BP.  
AC X35012;  
DT 01-JUL-1999 (first entry)  
DE Genomic DNA sequence of PCV strain 999PCV.  
KW PCV; strain PCV999; pig; PWS; porcine multisystemic wasting syndrome;  
OS Porcine circovirus  
PN FR2769321-A1.  
PD 09-APR-1999.  
PF 03-OCT-1997; 012382.  
PR 03-OCT-1997; FR-012382.  
PA (MERI-) MERIAL SAS.  
PA (UYBE-) UNIV QUEENS BELFAST.  
PA (UYSA-) UNIV SASKATCHEWAN.  
PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,  
PI Haines D, Harding J, Hassard L, Meehan B;  
DR WPI; 99-246947/21.  
PT New porcine circovirus from animals with porcine systemic wasting  
PT syndrome  
PS Claim 13; Fig 1; 35pp; French.  
CC The present sequence represents the genomic sequence of porcine  
CC circovirus (PCV) strain PCV999. The sequence is obtained from PCV  
CC isolated from a physiological or tissue sample, particularly from  
CC a lesion, from a pig showing symptoms of PWS (porcine multisystemic  
CC wasting syndrome), or cultured cells, infected with PCV isolated from  
CC such pigs. PCV (attenuated or inactivated), polypeptides derived from  
CC it, and vectors that express these polypeptides are all useful in  
CC vaccines, suitable for administration to adult or young pigs, or to  
CC pregnant sows (for passive immunization of their offspring). DNA  
CC isolated from PCV is used for in vivo or in vitro expression of viral  
CC polypeptides, also as probes or primers for diagnosis in usual  
CC hybridization or amplification assays. These polypeptides may also be  
CC used diagnostically to detect PCV-specific antibodies, while antibodies  
CC raised against the polypeptides can be used to detect antigens, in any  
CC usual immunoassay format.  
SQ Sequence 1768 BP; 453 A; 359 C; 486 G; 455 T;

alignment\_scores:  
Quality: 314.00 Length: 59  
Ratio: 5.322 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 98.305  
  
alignment\_block:  
US-09-209-961-20 x X35012 ..  
  
Align seg 1/1 to: X35012 from: 1 to: 1768  
  
1 MetTyrThrSerLeuTrpGlyHisLeuGlyValValLysAlaAsnGlyLe 17  
|||||  
901 ATGTACACGTCATTTGGGGCCACCTGGGTGGTAAAGCAATGGGCT 950  
|||||  
17 uLeuIleLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrS 34  
|||||  
951 GCTAATTTTGCAGACCCGGAACCACTACTGGAAACCACTAGAAACAA 1000  
|||||  
34 exGlyGlyMetValThrMetValLysLysTrpLeuLeuLeuMetThrPhe 50  
|||||  
1001 GTGGTGGGATGGTTACCATGGGGAAGAGTGGTGTATTGATGACTTTT 1050  
|||||  
51 MetAlaGlyCysArgGlyMetIleTyr 59  
|||||  
1051 ATGGCTGGTGGTGGGATGATCTAC 1077  
  
seq\_name: N\_Geneseq\_36:X35213

```

seq_documentation_block:
ID X35213 standard; DNA; 1768 BP.
AC X35213.
DT 01-JUL-1999 (first entry)
DE Nucleotide sequence of PCV isolate Impl010.
KW PCV isolate; type II porcine circovirus; PCV; PMWS;
OS Porcine multisystemic wasting syndrome; pig; vaccine; ss.
PN FR2769322-Al.
PD 09-APR-1999.
PF 22-JAN-1998; 000873.
PR 03-OCT-1997; FR-012382.
PA (MERI-) MERIAL SAS.
PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,
PI Haines D, Harding J, Hassard L, Meehan B;
DR WPI; 99-246948/21.
PT New type II porcine circovirus, used for, e.g. passive immunization
of pregnant sows
PS Claim 14; Fig 4; 48pp; French.
CC The present sequence represents the nucleotide sequence of PCV isolate
Impl010. The specification describes a preparation of type II
porcine circovirus (PCV), which is particularly isolated from a lesion,
from a pig with symptoms of PMWS (porcine multisystemic wasting
syndrome). PCV (attenuated or inactivated), polypeptides derived from
it, and vectors that express these polypeptides are all useful in
CC vaccines, suitable for administration to adult or young pigs, or to
pregnant sows (for passive immunization of their offspring). DNA
isolated from PCV is used for in vivo or in vitro expression of viral
polypeptides, also as probes or primers for diagnosis in usual
hybridization or amplification assays. These polypeptides may also be
used diagnostically to detect PCV-specific antibodies, while antibodies
raised against the polypeptides can be used to detect antigens, in any
usual immunoassay format.
SQ Sequence 1768 BP; 451 A; 361 C; 495 G; 461 T;

alignment_scores:
Quality: 314.00 Length: 59
Ratio: 5.322 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.305

alignment_block:
US-09-209-961-20 x X35213 ..
Align seg 1/1 to: X35213 from: 1 to: 1768

1 MetTyrThrSerLeuTrpGlyHisLeuGlyValValLysAlaAsnGlyLe 17
|||||
900 ATGTACACGTCATTGTGGGCCACCTGGTGTGTTAAAGCAAAATGGGCT 949

17 uLeuIleLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrs 34
|||||
950 GCTAATTTCAGACCCGGAACACACATCTGGAACACCACTAGAACAA 999

34 eRGlyGlyMetValThrMetValLysLysTrpLeuLeuLeuMetThrPhe 50
|||||
1000 GTGGTGGGATGGTTACCATGGTGAAGAGTGGTTGTTATTGATGACTTTT 1049

51 MetAlaGlyCysArgGlyMetIleTyr 59
|||||
1050 ATGGCTGGCTGCCGTGGGATGATCTAC 1076

seq_name: N_Geneseq_36:X35212

seq_documentation_block:
ID X35212 standard; DNA; 1768 BP.
AC X35212.
DT 01-JUL-1999 (first entry)
DE Nucleotide sequence of PCV isolate Imp999 (corrected version).
KW PCV isolate; type II porcine circovirus; PCV; PMWS;
OS Porcine multisystemic wasting syndrome; pig; vaccine; ss.
PN FR2769322-Al.
PD 09-APR-1999.
PF 22-JAN-1998; 000873.
PR 03-OCT-1997; FR-012382.
PA (MERI-) MERIAL SAS.
PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,
PI Haines D, Harding J, Hassard L, Meehan B;
DR WPI; 99-246948/21.
PT New type II porcine circovirus, used for, e.g. passive immunization
of pregnant sows
PS Claim 14; Fig 4; 48pp; French.
CC The present sequence represents the nucleotide sequence of PCV isolate
Impl010. The specification describes a preparation of type II
porcine circovirus (PCV), which is particularly isolated from a lesion,
from a pig with symptoms of PMWS (porcine multisystemic wasting
syndrome). PCV (attenuated or inactivated), polypeptides derived from
it, and vectors that express these polypeptides are all useful in
CC vaccines, suitable for administration to adult or young pigs, or to
pregnant sows (for passive immunization of their offspring). DNA
isolated from PCV is used for in vivo or in vitro expression of viral
polypeptides, also as probes or primers for diagnosis in usual
hybridization or amplification assays. These polypeptides may also be
used diagnostically to detect PCV-specific antibodies, while antibodies
raised against the polypeptides can be used to detect antigens, in any
usual immunoassay format.
SQ Sequence 1768 BP; 453 A; 360 C; 494 G; 461 T;

alignment_scores:
Quality: 314.00 Length: 59
Ratio: 5.322 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.305

alignment_block:
US-09-209-961-20 x X35212 ..
Align seg 1/1 to: X35212 from: 1 to: 1768

1 MetTyrThrSerLeuTrpGlyHisLeuGlyValValLysAlaAsnGlyLe 17
|||||
900 ATGTACACGTCATTGTGGGCCACCTGGTGTGTTAAAGCAAAATGGGCT 949

17 uLeuIleLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrs 34
|||||
950 GCTAATTTCAGACCCGGAACACACATCTGGAACACCACTAGAACAA 999

34 eRGlyGlyMetValThrMetValLysLysTrpLeuLeuLeuMetThrPhe 50
|||||
1000 GTGGTGGGATGGTTACCATGGTGAAGAGTGGTTGTTATTGATGACTTTT 1049

51 MetAlaGlyCysArgGlyMetIleTyr 59
|||||
1050 ATGGCTGGCTGCCGTGGGATGATCTAC 1076

seq_name: N_Geneseq_36:X35214

seq_documentation_block:
ID X35214 standard; DNA; 1768 BP.
AC X35214.
DT 01-JUL-1999 (first entry)
DE Nucleotide sequence of PCV isolate Imp999 (uncorrected version).
KW PCV isolate; type II porcine circovirus; PCV; PMWS;
OS Porcine multisystemic wasting syndrome; pig; vaccine; ss.
PN FR2769322-Al.
PD 09-APR-1999.
PF 22-JAN-1998; 000873.
PR 03-OCT-1997; FR-012382.
PA (MERI-) MERIAL SAS.
PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,
PI Haines D, Harding J, Hassard L, Meehan B;
DR WPI; 99-246948/21.
PT New type II porcine circovirus, used for, e.g. passive immunization
of pregnant sows
PS Claim 14; Fig 3; 48pp; French.
CC The present sequence represents the nucleotide sequence of PCV isolate
Impl999. The specification describes a preparation of type II
porcine circovirus (PCV), which is particularly isolated from a lesion,
from a pig with symptoms of PMWS (porcine multisystemic wasting
syndrome). PCV (attenuated or inactivated), polypeptides derived from
it, and vectors that express these polypeptides are all useful in
CC vaccines, suitable for administration to adult or young pigs, or to
pregnant sows (for passive immunization of their offspring). DNA
isolated from PCV is used for in vivo or in vitro expression of viral
polypeptides, also as probes or primers for diagnosis in usual
hybridization or amplification assays. These polypeptides may also be
used diagnostically to detect PCV-specific antibodies, while antibodies
raised against the polypeptides can be used to detect antigens, in any
usual immunoassay format.
SQ Sequence 1768 BP; 453 A; 360 C; 494 G; 461 T;

alignment_scores:
Quality: 314.00 Length: 59
Ratio: 5.322 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.305

alignment_block:
US-09-209-961-20 x X35212 ..
Align seg 1/1 to: X35212 from: 1 to: 1768

1 MetTyrThrSerLeuTrpGlyHisLeuGlyValValLysAlaAsnGlyLe 17
|||||
900 ATGTACACGTCATTGTGGGCCACCTGGTGTGTTAAAGCAAAATGGGCT 949

17 uLeuIleLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrs 34
|||||
950 GCTAATTTCAGACCCGGAACACACATCTGGAACACCACTAGAACAA 999

34 eRGlyGlyMetValThrMetValLysLysTrpLeuLeuLeuMetThrPhe 50
|||||
1000 GTGGTGGGATGGTTACCATGGTGAAGAGTGGTTGTTATTGATGACTTTT 1049

51 MetAlaGlyCysArgGlyMetIleTyr 59
|||||
1050 ATGGCTGGCTGCCGTGGGATGATCTAC 1076

seq_name: N_Geneseq_36:X35212

```

PS Claim 14; Fig 6; 48pp; French.  
 CC The present sequence represents the nucleotide sequence of PCV isolate  
 CC Imp999. The specification describes a preparation of type II  
 CC porcine circovirus (PCV), which is particularly isolated from a lesion,  
 CC from a pig with symptoms of PMWS (porcine multisystemic wasting  
 CC syndrome). PCV (attenuated or inactivated), polypeptides derived from  
 CC it, and vectors that express these polypeptides are all useful in  
 CC vaccines, suitable for administration to adult or young pigs, or to  
 CC pregnant sows (for passive immunization of their offspring). DNA  
 CC isolated from PCV is used for in vivo or in vitro expression of viral  
 CC polypeptides, also as probes or primers for diagnosis in usual  
 CC hybridization or amplification assays. These polypeptides may also be  
 CC used diagnostically to detect PCV-specific antibodies, while antibodies  
 CC raised against the polypeptides can be used to detect antigens, in any  
 CC usual immunoassay format.  
 SQ Sequence 1768 BP; 453 A; 359 C; 486 G; 455 T;

alignment\_scores:  
 Quality: 314.00 Length: 59  
 Ratio: 5.322 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 98.305

alignment\_block:  
 US-09-209-961-20 x X35214 ..

Align seg 1/1 to: X35214 from: 1 to: 1768

1 MetTyrThrSerLeuTrpGlyHisLeuGlyValValLysAlaAsnGlyLe 17  
 |||||||||||||||||||||||||||||||||||||||||||||||||||  
 901 ATGTACACGTCATTTGGGGCCACCTGGGTGGTAAAGCAATGGGCT 950

17 uLeuLeuGlnThrArgLysProHisThrClyAsnHisLeuLysThrS 34  
 |||||||||||||||||||||||||||||||||||||||||||||||||||  
 951 GCTAATTTGCAGACCCCGGAACCACTACTGGAAACCACTAGAACAA 1000

34 eRGlyGlyMetValThrMetValLysLysTrpLeuLeuMetThrPhe 50  
 |||||||||||||||||||||||||||||||||||||||||||||||||||  
 1001 GTGGTGGGATGGTTACCATGGTGAGAGAGTGGTTGTTATTGATGACTTTT 1050

51 MetAlaGlyCysArgGlyMetIleTyr 59  
 |||||||||||||||||||||||||||||||||||||||||||||||||||  
 1051 ATGGCTGGCTGCCGGGGATGATCTAC 1077

seq\_name: N\_Geneseq\_36:X35210

seq\_documentation\_block:  
 ID X35210 standard; DNA; 1767 BP.  
 AC X35210;  
 DT 01-JUL-1999 (first entry)  
 DE Nucleotide sequence of PCV isolate Imp1011-48121.  
 KW PCV isolate; type II porcine circovirus; PCV; PMWS;  
 KW porcine multisystemic wasting syndrome; pig; vaccine; ss.  
 OS Porcine circovirus.  
 PN FR2769322-A1.  
 PD 03-APR-1999.  
 PF 22-JAN-1998; 000873.  
 PR 03-OCT-1997; FR-012382.  
 PA (MERI-) MERIAL SAS.  
 PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,  
 PI Haines D, Harding J, Hassard L, Meehan B;  
 DR WPI; 99-246948/21.  
 PT New type II porcine circovirus, used for, e.g. passive immunization  
 of pregnant sows  
 PS Claim 14; Fig 1; 48pp; French.  
 CC The present sequence represents the nucleotide sequence of PCV isolate  
 CC Imp1011-48121. The specification describes a preparation of type II  
 CC porcine circovirus (PCV), which is particularly isolated from a lesion,  
 CC from a pig with symptoms of PMWS (porcine multisystemic wasting  
 CC syndrome). PCV (attenuated or inactivated), polypeptides derived from  
 CC it, and vectors that express these polypeptides are all useful in  
 CC vaccines, suitable for administration to adult or young pigs, or to  
 CC pregnant sows (for passive immunization of their offspring). DNA

CC isolated from PCV is used for in vivo or in vitro expression of viral  
 CC polypeptides, also as probes or primers for diagnosis in usual  
 CC hybridization or amplification assays. These polypeptides may also be  
 CC used diagnostically to detect PCV-specific antibodies, while antibodies  
 CC raised against the polypeptides can be used to detect antigens, in any  
 CC usual immunoassay format.  
 SQ Sequence 1767 BP; 447 A; 360 C; 502 G; 458 T;

alignment\_scores:  
 Quality: 307.00 Length: 59  
 Ratio: 5.293 Gaps: 0  
 Percent Similarity: 98.305 Percent Identity: 96.610

alignment\_block:  
 US-09-209-961-20 x X35210 ..

Align seg 1/1 to: X35210 from: 1 to: 1767

1 MetTyrThrSerLeuTrpGlyHisLeuGlyValValLysAlaAsnGlyLe 17  
 |||||||||||||||||||||||||||||||||||||||||||||||||||  
 900 ATGTACACGTCATTTGGGGCCACCTGGGTGGTAAAGCAATGGGCT 949

17 uLeuLeuGlnThrArgLysProHisThrClyAsnHisLeuLysThrS 34  
 |||||||||||||||||||||||||||||||||||||||||||||||||||  
 950 GCTAATTTGCAGACCCCGGAACCACTACTGGAAACCACTAGAACAA 999

34 eRGlyGlyMetValThrMetValLysLysTrpLeuLeuMetThrPhe 50  
 |||||||||||||||||||||||||||||||||||||||||||||||||||  
 1000 GTGGTGGGATGGTTACCATGGTGAGAGTGGTTGTTATTGATGACTTTT 1049

51 MetAlaGlyCysArgGlyMetIleTyr 59  
 |||||||||||||||||||||||||||||||||||||||||||||||||||  
 1050 ATGGCTGGCTGCCCTGGGATGATCTAC 1076

seq\_name: N\_Geneseq\_36:X35211

seq\_documentation\_block:  
 ID X35211 standard; DNA; 1767 BP.  
 AC X35211;  
 DT 01-JUL-1999 (first entry)  
 DE Nucleotide sequence of PCV isolate Imp1011-48285.  
 KW PCV isolate; type II porcine circovirus; PCV; PMWS;  
 KW porcine multisystemic wasting syndrome; pig; vaccine; ss.  
 OS Porcine circovirus.  
 PN FR2769322-A1.  
 PD 09-APR-1999.  
 PF 22-JAN-1998; 000873.  
 PR 03-OCT-1997; FR-012382.  
 PA (MERI-) MERIAL SAS.  
 PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,  
 PI Haines D, Harding J, Hassard L, Meehan B;  
 DR WPI; 99-246948/21.  
 PT New type II porcine circovirus, used for, e.g. passive immunization  
 of pregnant sows  
 PS Claim 14; Fig 2; 48pp; French.  
 CC The present sequence represents the nucleotide sequence of PCV isolate  
 CC Imp1011-48285. The specification describes a preparation of type II  
 CC porcine circovirus (PCV), which is particularly isolated from a lesion,  
 CC from a pig with symptoms of PMWS (porcine multisystemic wasting  
 CC syndrome). PCV (attenuated or inactivated), polypeptides derived from  
 CC it, and vectors that express these polypeptides are all useful in  
 CC vaccines, suitable for administration to adult or young pigs, or to  
 CC pregnant sows (for passive immunization of their offspring). DNA  
 CC isolated from PCV is used for in vivo or in vitro expression of viral  
 CC polypeptides, also as probes or primers for diagnosis in usual  
 CC hybridization or amplification assays. These polypeptides may also be  
 CC used diagnostically to detect PCV-specific antibodies, while antibodies  
 CC raised against the polypeptides can be used to detect antigens, in any  
 CC usual immunoassay format.  
 SQ Sequence 1767 BP; 448 A; 359 C; 500 G; 460 T;

alignment\_scores:  
 Quality: 307.00 Length: 59  
 Ratio: 5.293 Gaps: 0  
 Percent Similarity: 98.305 Percent Identity: 96.610

alignment\_block:  
 US-09-209-961-20 x X35211 ..

Align seg 1/1 to: X35211 from: 1 to: 1767

```

1 MetTyrThrSerLeuTrpGlyHisLeuGlyValValLysAlaAsnGlyLe 17
|||||
900 ATGTACACGTCATGTGGGCCACCTGGGTGTGTAAGCAAAATGGGCT 949
|||||
17 uLeuLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrS 34
|||||
950 GCTAATTTTCAGACCCGGAACCACTACCTGGAACCACTAGAACAA 999
|||||
34 eRgLYGlyMetValThrMetValLysLysTrpLeuLeuMetThrPhe 50
|||||
1000 GTGGTGGGATGTTACCATGCTGAAGAGTGTGTTATTGATGACITTT 1049
|||||
51 MetAlaGlyCysArgGlyMetIleTyr 59
|||||
1050 ATGGCTGGCTGCCCTGGGATGATCTAC 1076

```

seq\_name: N\_Geneseq\_36.X35013

seq\_documentation\_block:  
 ID X35013 standard; DNA; 1759 BP.  
 AC X35013;  
 DT 01-JUL-1999 (first entry)  
 KW Nucleotide sequence which has homology to PCV sequence.  
 DE PCV; strain PCV999; pig; PMWS; porcine multisystemic wasting syndrome;  
 KW vaccine; ss.  
 OS Sus sp.  
 PN FR2769321-AL.  
 PD 09-APR-1999.  
 PF 03-OCT-1997; 012382.  
 PR 03-OCT-1997; FR-012382.  
 PA (MERI-) MERIAL SAS.  
 PA (UYBE-) UNIV QUEENS BELFAST.  
 PA (UYSA-) UNIV SASKATCHEWAN.  
 PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,  
 PI Haines D, Harding J, Hassard L, Meehan B;  
 DR WPI; 99-246947/21.  
 PT New porcine circovirus from animals with porcine systemic wasting syndrome  
 PS Disclosure: Fig 2; 35pp; French.  
 CC The specification describes a genomic sequence of porcine  
 CC circovirus (PCV) strain PCV999. The sequence is obtained from PCV  
 CC isolated from a physiological or tissue sample, particularly from  
 CC a lesion, from a pig showing symptoms of PMWS (porcine multisystemic  
 CC wasting syndrome), or cultured cells, infected with PCV isolated from  
 CC such pigs. PCV (attenuated or inactivated), polypeptides derived from  
 CC it, and vectors that express these polypeptides are all useful in  
 CC vaccines, suitable for administration to adult or young pigs, or to  
 CC pregnant sows (for passive immunization of their offspring). DNA  
 CC isolated from PCV is used for in vivo or in vitro expression of viral  
 CC polypeptides, also as probes or primers for diagnosis in usual  
 CC hybridization or amplification assays. These polypeptides may also be  
 CC used diagnostically to detect PCV-specific antibodies, while antibodies  
 CC raised against the polypeptides can be used to detect antigens, in any  
 CC usual immunoassay format. The present sequence appears in the  
 CC specification.  
 SQ Sequence 1759 BP; 475 A; 510 C; 342 G; 432 T;

alignment\_scores:  
 Quality: 114.00 Length: 59  
 Ratio: 2.923 Gaps: 0  
 Percent Similarity: 66.102 Percent Identity: 47.458

alignment\_block:  
 US-09-209-961-20 x X35013/rev ..

Align seg 1/1 to reverse of: X35013 from: 1 to: 1759

```

1 MetTyrThrSerLeuTrpGlyHisLeuGlyValValLysAlaAsnGlyLe 17
|||||
872 CTGTACACGTCATGTGGGCCCGCCGGTGTGGAAGAGCCAGTGGGCC 823
|||||
17 uLeuLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrS 34
|||||
822 CGTAATTTTGTGAGCCCTAGGACACCTACTGGAAGCCCTAGTAATAA 773
|||||
34 eRgLYGlyMetValThrMetValLysLysTrpLeuLeuMetThrPhe 50
|||||
772 GTGGTGGGATGATCATGGAAGAAGTGTGTTTGGATGATTTT 723
|||||
51 MetAlaGlyCysArgGlyMetIleTyr 59
|||||
722 ATGGCTGGTTACCTGGGATGATCTAC 696

```

seq\_name: N\_Geneseq\_36.X41449

seq\_documentation\_block:  
 ID X41449 standard; cDNA; 315 BP.  
 AC X41449;  
 DT 22-JUN-1999 (first entry)  
 DE Human secreted protein 5; EST SEQ ID NO: 108 from WO 9906553.  
 KW Human; secreted protein; EST; expressed sequence tag; diagnosis;  
 KW forensic; gene therapy; chromosome mapping; signal peptide;  
 KW upstream regulatory sequence; cytokine activity; cell proliferation;  
 KW differentiation; haematopoiesis regulation; tissue growth regulation;  
 KW reproductively hormone regulation; chemotactic; chemokinetic; haemostatic;  
 KW thrombolytic; antiinflammatory; tumour inhibition; antitumour; ds.  
 OS Homo sapiens.  
 PN WO9906553-A2.  
 PD 11-FEB-1999.  
 PF 31-JUL-1998; IB1237.  
 PR 01-AUG-1997; US-905051.  
 PA (GEST) GENSET.  
 PI Duclert A, Dumas Milne Edwards J, Lacroix B;  
 DR WPI; 99-153783/13.  
 DR P-PSDB; Y12591.  
 PT New nucleic acids encoding human secreted proteins - obtained from  
 PT cDNA libraries derived from umbilical cord, lymph ganglia,  
 PT lymphocytes and placental tissue  
 PS Claim 1; Page 219-220; 41pp; English.  
 CC X41379 to X41526 represent 5' expressed sequence tags (ESTs) for human  
 CC secreted proteins, and encode the proteins given in Y12521 to Y12668,  
 CC respectively. The proteins given represent the signal peptide and an  
 CC N-terminal fragment of a secreted protein. The nucleic acid sequences  
 CC can be used for producing secreted human gene products. They can also  
 CC be used to develop products for diagnosis and therapy. The proteins  
 CC obtained may have cytokine activity, cell proliferation/differentiation  
 CC activity, haematopoiesis regulating activity, tissue growth regulating  
 CC activity, reproductively hormone regulating activity, chemotactic/  
 CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/  
 CC ligand activity, antiinflammatory activity, tumour inhibition activity  
 CC or other activities. The products can be used in forensic, gene therapy  
 CC and chromosome mapping procedures. The sequences can also be used for  
 CC obtaining corresponding promoter sequences. The nucleic acids encoding  
 CC the signal peptide can be used for directing extracellular secretion of  
 CC a polypeptide or the insertion of a polypeptide into a membrane, or  
 CC importing a polypeptide into a cell.  
 SQ Sequence 315 BP; 67 A; 85 G; 71 T;

alignment\_scores:  
 Quality: 65.00 Length: 42  
 Ratio: 2.500 Gaps: 1  
 Percent Similarity: 61.905 Percent Identity: 35.714

alignment\_block:

```

US-09-209-961-20 x X41449/rev ..
Align seg 1/1 to reverse of: X41449 from: 1 to: 315

      7 GlyHisLeuGlyValValLysAlaAsnGlyLeuLeuLeuLeuGlnThrAr 23
      ||||| ||||| :||| :||| ||| |||||
289 GGCCACTGGGACTGTCCTCTGGAGGCCATCTGTGCGCCGACGCG 240
      ||||| ||||| :||| :||| ||| |||||
23 gLysProHisThrGlyAsnHis.....LeuLysThrSerGlyGlyMetV 38
      ||||| ||||| :||| :||| ||| |||||
239 CCGTCCATGGTCAGGACACACCCAGCTGASTGACGCTGAGGGGCTACAG 190
      ||||| ||||| :||| :||| ||| |||||
38 alThrMetValLysLysTrpLeuLeu 46
      ||||| ||||| :||| :||| ||| |||||
189 TGAACCTGGACAGCTGGCTTCG 164

seq_name: N_Geneseq_36:Q60652

seq_documentation_block:
ID Q60652 standard; cDNA; 379 BP.
AC Q60652;
DT 16-MAR-1994 (first entry)
DE Human brain Expressed Sequence Tag EST02666.
KW Gene transcription product; genetic markers; tagging; in vivo;
OS Homo sapiens.
PN WO9316178-A.
PD 19-AUG-1993.
PF 12-FEB-1993; U01294.
PR 12-FEB-1992; US-837195.
PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI Adams MD, Moreno RF, Venter CJ;
DR WPI: 93-272882/34.
PT Enriched oligonucleotides and corresp. sequences - used as
PT markers for human genes transcribed in-vivo, facilitate tagging
PT of most human genes
PS Example 4; Page 368; 500pp; English.
CC The Expressed Sequence Tag was isolated from a human brain cDNA
CC library as part of a large set of ESTs which can be used as markers
CC for human genes transcribed in vivo. They can be used to facilitate
CC tagging of most human genes, for mapping locations of expressed genes
CC on chromosomes, for individual or forensic identification, for mapping
CC locations of disease-associated genes, for identification of tissue
CC type, and for prep. of antisense sequences, probes and constructs.
CC EST02666 has a "poor" coding probability as evaluated using the
CC coding-region prediction program CRM. See also Q59041-Q61440.
SQ Sequence 379 BP; 70 A; 117 C; 76 G; 110 T;

alignment_scores:
Quality: 57.50 Length: 36
Ratio: 2.614 Gaps: 1
Percent Similarity: 61.111 Percent Identity: 38.889

alignment_block:
US-09-209-961-20 x Q60652 ..
Align seg 1/1 to: Q60652 from: 1 to: 379

17 LeuLeuLeuLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysTh 33
||||| ||||| ||||| :||| :||| ||| |||||
38 CTCCTTCGCTCCAGCCACATCACCAGCCGCGGAGTTCCTCCACCTAGCACC 87
||||| ||||| ||||| :||| :||| ||| |||||
33 rSerGlyGlyMetValThrMetValLysLysTrpLeuLeuLeuMetThrP 50
||||| ||||| ||||| :||| :||| ||| |||||
88 TTCCTCGGAATNATCTCC.....CCCTGGTGGCTCTTTCTACTT 128
||||| ||||| ||||| :||| :||| ||| |||||
50 heMetAla 52
:: |||
129 ATTCAGCC 136

seq_name: N_Geneseq_36:X40573

seq_documentation_block:
ID X40573 standard; cDNA; 168 BP.
AC X40573;
DT 18-JUN-1999 (first entry)
DE Human secreted protein 5, EST SEQ ID No: 173.
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide; prostate;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.
OS Homo sapiens.
PN WO9906550-A2.
PD 11-FEB-1999.
PF 31-JUL-1998; IB1232.
PR 01-AUG-1997; US-905144.
PA (GEST ) GENSET.
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
DR WPI: 99-153780/13.
DR P-PSDB: Y11851.
PT New isolated prostate-derived nucleic acids - used to develop
PT products which may have cytokine, immune regulatory, haematopoiesis
PT regulating, anti-inflammatory or tumour inhibition activity
PS Claim 1; Page 276; 675pp; English.
CC X40438 to X40715 represent 5' expressed sequence tags (ESTs) for human
CC secreted proteins expressed in prostate, and encode the proteins given in
CC peptide and an N-terminal fragment of a secreted protein. The nucleic
CC acid sequences can be used for producing secreted human gene products.
CC They can also be used to develop products for diagnosis and therapy. The
CC proteins obtained may have cytokine activity, cell proliferation and
CC differentiation activity, haematopoiesis regulating activity, tissue
CC growth regulating activity, reproductive hormone regulating activity,
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
CC receptor/ligand activity, anti-inflammatory activity, tumour inhibition
CC activity or other activities. The products can be used in forensic, gene
CC therapy and chromosome mapping procedures. The sequences can also be used
CC for obtaining corresponding promoter sequences. The nucleic acids
CC encoding the signal peptides can be used for directing extracellular
CC secretion of a polypeptide or the insertion of a polypeptide into a
CC membrane, or importing a polypeptide into a cell.
SQ Sequence 168 BP; 61 A; 27 C; 39 G; 39 T;

alignment_scores:
Quality: 54.00 Length: 33
Ratio: 2.455 Gaps: 0
Percent Similarity: 66.667 Percent Identity: 33.333

alignment_block:
US-09-209-961-20 x X40573/rev ..
Align seg 1/1 to reverse of: X40573 from: 1 to: 168

6 TrpGlyHisLeuGlyValValLysAlaAsnGlyLeuLeuLeuGlnTh 22
||||| ||||| ||||| :||| :||| ||| |||||
102 TGGCAGCATCTTCAATACTGTAATCTGTCTGGCTCTCTGAAGACGATGAT 53
||||| ||||| ||||| :||| :||| ||| |||||
22 rArgLysProHisThrGlyAsnHisLeuLysThrSerGlyGlyMetVal 38
||||| ||||| ||||| :||| :||| ||| |||||
52 GAGGAAGAGGAACACATAGGATCATCTGAAGAAAAGACGCTCACTGATC 4

seq_name: N_Geneseq_36:Q89836

seq_documentation_block:
ID Q89836 standard; cDNA; 309 BP.
AC Q89836;
DT 04-DEC-1995 (first entry)
DE Human death associated protein DAP-1.
KW Death associated protein; DAP; cytokine; cell death; ss.
OS Homo sapiens.
PF Key Location/Qualifiers
FT CDS 1..309

```

```

FT W09510630-A. /*tag= a
PN 20-APR-1995.
PD 12-OCT-1994; U11598.
PF 12-OCT-1993; IL-107250.
PR 12-OCT-1993; IL-107250.
PA (RYCU) RYCUS A.
PA (YEDA) YEDA RES & DEV CO LTD.
PI Kimchi A.
DR WPI; 95-178528/23.
DR P-PSDB; R4204.
PT DNA whose expression mediates cytokine-induced programmed cell
PT death - used to treat diseases or disorders associated with
PT uncontrolled, pathological cell growth or cytokine-induced
PT programmed cell death.
PS Claim 2; Fig 6; 6lpp; English.
CC DAP genes seem to play an imp. role in programmed cell death and the
CC inhibition of their expression protects the cell from cytokine-
CC promoted cell death. An HL-60 cDNA library constructed in lambda
CC gt10 vector was screened with the cDNA insert of pTK01-230. Two
CC independent clones, lambda2 and lambda2, almost completely
CC overlapping and carrying cDNA inserts of about 2.3 Kb were analysed.
CC Lambda2 cDNA clone encompasses the 5'-UTR, short coding region(s)
CC and a relatively long 3' UTR that constitutes more than 60% of the
CC cDNA clone. The nt. sequence of the cDNA carried by lambda2 and
CC predicted AA pattern are present in Q89835/R4204. The ORF is
CC preceded by an extremely GC-rich 5' UTR and potentially codes for
CC a protein consisting of 102 AAs with calculated MW of 11.2 kDa.
CC It has an isoelectric pt. of 10 and is rich in prolines. Q89836
CC is the DAP-1 ORF.
SQ Sequence 309 BP; 90 A; 96 C; 79 G; 44 T;

alignment_scores:
Quality: 55.50 Length: 23
Ratio: 3.083 Gaps: 1
Percent Similarity: 78.261 Percent Identity: 52.174

alignment_block:
US-09-209-961-20 x Q89836 ..
Align seg 1/1 to: Q89836 from: 1 to: 309

7 GlyHisLeuGlyValValLysAlaAsnGlyLeuLeuLeuLeuGlnThrAr 23
||||| :|||:|||||:|||||:|||||:|||||:
40 GGACACCCGCCCGCTGAAGCTGGTGAATGCGAATGTCGAG...AA 86

23 GlyProHisThrGlyAsn 29
: |||||:|||||:
87 ACACCCACATACAGGAGAC 105

seq_name: N_Geneseq_36:Q46896

seq_documentation_block:
ID Q46896 standard; DNA; 114 BP.
AC Q46896.
DE 14-FEB-1994 (first entry)
DE JES1-39D10 primer B2137CC.
KW Heavy; light; chain; anti-human; monoclonal; antibody; JES1-39D10;
KW complementarity determining region; CDR; loop analysis; IL-5; human;
KW Kabat determination; grafting; PCR; polymerase chain reaction; amplify;
KW primer; ss.
OS Synthetic.
PN W09316184-A.
PD 19-AUG-1993.
PF 04-FEB-1993; U00759.
PR 06-FEB-1992; US-832842.
PA (SCHE) SCHERING CORP.
PI Abrams JS, Chou C, Jenh C, Murgolo NJ, Petro ME;
PI Silver JE, Tindall S, Windsor WT, Zavodny PJ;
DR WPI; 93-272888/34.
PT Humanised monoclonal antibody - comprises variable animal region
PT and constant human region, binds to human interleukin-5
PS Example; Page 71; 118pp; English.

```

```

CC The sequences given in Q46892-99 are primers which were used in the
CC amplification of the heavy and light chains of the anti-human mono-
CC clonal antibody JES1-39D10. These primer sequences were based on
CC the Ig2a/kappa isotype of JES1-39D10. The amplification products
CC were isolated by gel electrophoresis and cloned separately into pUC19.
CC The ligated plasmids were used to transform E. coli DH5-alpha. The
CC position of the complementarity determining regions (CDRs) in the
CC amplified and expressed sequences, was determined by the method of
CC Kabat et al and by loop analysis. These CDRs may be used to produce
CC humanised antibodies by grafting them onto a human antibody. Amino
CC acid residues from outside the CDR may also be included in the graft
CC if they interact with the CDRs or with IL-5.
SQ Sequence 114 BP; 27 A; 21 C; 33 G; 33 T;

alignment_scores:
Quality: 51.00 Length: 20
Ratio: 3.643 Gaps: 0
Percent Similarity: 70.000 Percent Identity: 55.000

alignment_block:
US-09-209-961-20 x Q46896/rev ..
Align seg 1/1 to reverse of: Q46896 from: 1 to: 114

17 LeuLeuLeuLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysTh 33
||||| :|||:|||||:|||||:|||||:
105 CTCCTCCAGATTCACACAGTTTCACCTCACACTGGATACCATTTAAAGT 56

33 rSerGlyGly 36
: |||
55 GTTAAAGGA 46

seq_name: N_Geneseq_36:T31291

seq_documentation_block:
ID T31291 standard; cDNA; 3269 BP.
AC T31291.
DE 24-FEB-1997 (first entry)
DE Rat poly-immunoglobulin receptor, cDNA.
KW Rat; immunoglobulin; receptor; protection protein; mutants;
KW heavy chain; antigen binding domain; protection; pathogen;
KW mucosal; environment; gastrointestinal; passive; immunisation;
KW Guy's 13 antibody; prevention; dental caries; Streptococcus;
KW poly; sorbinus; ss.
OS Rattus rattus.
FH Key
FT Location/Qualifiers
FT cds
FT 74..2383
FT /*tag= a
FT
PN W09621012-A1.
PD 11-JUL-1996.
PF 27-DEC-1995; U16889.
PR 30-DEC-1994; US-367395.
PR 04-MAY-1995; US-434000.
PA (PLAN-) PLANT BIOTECHNOLOGY INC.
PA (UNNE-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.
PA (PLAN-) PLANET BIOTECHNOLOGY INC.
PI Hiatt AC, Lehner T, Ma JKC;
DR WPI; 96-333987/33.
DR P-PSDB; W03181.
PT Immunoglobulin and protection protein complex and its prodn. in
PT plants - useful for passive immunisation against mucosal antigens,
PT esp. against S. mutans and S. sorbinus to prevent dental caries
PS Disclosure; Pages 123-127; 152pp; English.
CC The present sequence encodes the rat poly-immunoglobulin (Ig)
CC receptor, a portion of which corres. to residues 1-627, pref.
CC 1-606, or esp. residues 13-45, 1-120, 110-230, 210-340, 320-450,
CC 440-550, 550-606 or 550-627 comprises a protection protein (PP).
CC The Ig of the invention comprises a pp as above in association with
CC an Ig derived heavy chain, having at least a portion of an antigen
CC (Ag) binding domain. The pp protects the Ig in harsh mucosal, e.g.
CC gastrointestinal, environments, therefore enhancing its
CC effectiveness in passively immunising animals against mucosal

```

**alignment\_block:**





OM of: US-09-209-961-20 to: Issued\_Patents\_NA:\* out\_format : pfs  
Date: Dec 31, 1999 12:26 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:  
-MODEL=framet.p2n.model -DEV=xlp  
-O/cgn2\_1/USPTO.spool/US0209961/runat.22121999.101043.17921/app\_query.fasta.1  
-DB=issued\_Patents\_NA -QFMT=fastap -SUFFIX=rni -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -QGAPEXT=0.500 -XGAPOP=10.000 -XGAPEXT=0.500  
-FCAPOP=6.000 -FCAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62  
-TRANS=human40 cdi -LIST=45 -DOALIGN=200 -THR\_SCORE=escore  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US0209961  
-NCPU=6 -ICPU=3 -NO\_XLPXY -WAIT -THREADS=1

Search information block:  
Query: US-09-209-961-20  
Query length: 59  
Database: Issued\_Patents\_NA.\*  
Database sequences: 195662  
Database length: 53930012  
Search time (sec): 88.010000

score_list:	Sequence	Strd Orig	Zscore	Escore	Len	Documentation
	/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-533-298-3		58.00	131.84	6.02	980
	/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-222-177A-13		50.00	127.97	9.89	190
	/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-354-404-21		56.50	123.66	17.19	1408
	/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-327-451E-21		56.50	123.66	17.19	1408
	/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-148-910-10		44.00	122.33	20.39	71
	/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-448-937A-10		44.00	122.33	20.39	71
	/cgn2_6/ptodata/1/ina/backfiles1.seq:5218099-4		49.00	121.29	23.28	270
	/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-470-892-4		51.00	121.06	23.97	453
	/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-050-073-2		48.50	119.95	27.66	269
	/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-050-073-9		48.50	119.95	27.66	269
	/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-050-073-10		48.50	119.95	27.66	269
	/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-050-073-12		48.50	119.95	27.66	269
	/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-264-250A-2		47.50	118.40	33.71	241
	/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-050-073-8		47.50	117.18	39.44	269
	/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-602-723-33		53.00	116.69	42.00	1103
	/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-853-217-26		48.00	116.49	43.07	324
	/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-158-682A-3		52.00	116.45	43.31	879
	/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-721-488-12		48.00	116.39	43.64	327
	/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-318-970B-42		49.00	115.90	46.45	438
	/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-050-073-16		46.50	114.41	56.24	269
	/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-050-073-17		46.50	114.41	56.24	269
	/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-050-073-32		46.50	114.41	56.24	269
	/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-887-798-36		37.00	114.14	58.24	26
	/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-07-953-230A-6		50.50	113.38	64.24	798
	/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-039-137-16		45.50	111.69	79.77	268
	/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-449-287-3		47.00	110.84	88.96	420
	/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-980-060-14		47.50	110.76	89.88	479
	/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-403-762A-81		37.00	109.07	111.60	41
	/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-050-073-24		44.50	108.88	114.35	269
	/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-050-073-35		44.50	108.88	114.35	269
	/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-726-306A-28		48.00	108.57	118.95	660
	/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-050-073-16		46.50	108.23	124.36	771
	/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-050-073-17		46.50	108.23	124.36	771
	/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-801-164A-5		41.50	107.69	133.17	142
	/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-801-164A-6		41.50	107.69	133.17	142
	/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-602-359A-26		48.00	107.06	144.40	756
	/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-533-298-3		53.50	106.58	153.50	3
	/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-841-662-13		45.50	106.50	155.10	427

/cgn2\_6/ptodata/1/ina/5A\_COMB.seq:US-08-209-797-13 - 45.50 106.50 155.10 427  
/cgn2\_6/ptodata/1/ina/5C\_COMB.seq:US-08-669-685-13 - 45.50 106.50 155.10 427  
/cgn2\_6/ptodata/1/ina/5A\_COMB.seq:US-08-533-298-3  
seq\_name: /cgn2\_6/ptodata/1/ina/5C\_COMB.seq:US-08-533-298-3  
seq\_documentation\_block:  
: Sequence 3, Application US/08533298  
: Patent No. 5851791  
: GENERAL INFORMATION:  
: APPLICANT: Vierstra, Richard D  
: APPLICANT: Gosink, Mark M  
: TITLE OF INVENTION: Ubiquitin Conjugating Enzyme (E2) Fusion  
: TITLE OF INVENTION: Proteins  
: NUMBER OF SEQUENCES: 18  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Quarles and Brady  
: STREET: 1 South Pinckney Street - Suite 600  
: CITY: Madison  
: STATE: WI  
: COUNTRY: USA  
: ZIP: 53703  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/533,298  
: FILING DATE: 25-SEP-1995  
: CLASSIFICATION: 536  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: 08/070,157  
: FILING DATE: 28-MAY-1993  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Seay, Nicholas J  
: REGISTRATION NUMBER: 27,386  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: 608-251-5000  
: TELEFAX: 608-251-9166  
: INFORMATION FOR SEQ ID NO: 3:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 980 base pairs  
: TYPE: nucleic acid  
: STRANDEDNESS: double  
: TOPOLOGY: linear  
: MOLECULE TYPE: cDNA  
: ORIGINAL SOURCE:  
: ORGANISM: Triticum vulgare  
: FEATURE:  
: NAME/KEY: CDS  
: LOCATION: 60..614  
: PUBLICATION INFORMATION:  
: AUTHORS: Sullivan, M L  
: AUTHORS: Vierstra, R D  
: JOURNAL: Proc. Natl. Acad. Sci. U.S.A.  
: VOLUME: 86  
: PAGES: 9861-9865  
: DATE: 1989  
: US-08-533-298-3  
alignment\_scores:  
Quality: 58.00 Length: 57  
Ratio: 1.758 Gaps: 4  
Percent Similarity: 57.895 Percent Identity: 31.579  
alignment\_block:  
US-09-209-961-20 x US-08-533-298-3/rev ..  
Align seq 1/1 to reverse of: US-08-533-298-3 from: 1 to: 980  
6 TrpGlyHis.....LeuGlyValValLysAlaAsnGlyLeuLeuI 19







```
; APPLICATION NUMBER: 350,570; 334,701; 228,334; 215,728; 846,757
; FILING DATE: 11-MAY-1989
; APPLICATION NUMBER: 334,701
; FILING DATE: 06-APR-1989
; APPLICATION NUMBER: 228,334
; FILING DATE: 04-AUG-1988
; APPLICATION NUMBER: 215,728
; FILING DATE: 06-JUL-1988
; APPLICATION NUMBER: 846,757
; FILING DATE: 01-APR-1986
; SEQ ID NO:4
; LENGTH: 270
5218099-4
```

```
alignment_scores:
  Quality: 49.00      Length: 37
  Ratio: 1.885       Gaps: 1
Percent Similarity: 70.270 Percent Identity: 29.730

alignment_block:
US-09-209-961-20 x 5218099-4/rev ..
Align seg 1/1 to reverse of: 5218099-4 from: 1 to: 270

17 LeuLeuLeuLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysTh 33
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
130 TTACTTTATGTAGCGGAATGCGCATCTTGCTCACAATAGCGGTC 81
: : : : : : : : : : : : : : : : : : : : :
33 rserGlyGlyMetValThrMetValLysLysTrpLeuLeuLeuMetThr 50
: : : : : : : : : : : : : : : : : : : : :
80 GAAG.....CAATCGAGCTGTGCGCGCATATTTGCCACTTTTGT 37

50 heMetAlaGly 53
||||| :|||
36 TTCAGTTGGC 26
```

seq\_name: /cgn2\_5/ptodata/1/ina/5B\_COMB.seq:US-08-470-892-4

```
seq_documentation_block:
; Sequence 4, Application US/08470892
; Patent No. 5707839
; GENERAL INFORMATION:
; APPLICANT: DENOYA, CLAUDIO D.
; APPLICANT: STUTZMAN-ENGWALL, KIM J.
; TITLE OF INVENTION: GENES ENCODING BRANCHED-CHAIN
; TITLE OF INVENTION: ALPHA-KETOACID DEHYDROGENASE COMPLEX FROM STREPTOMYCES
; TITLE OF INVENTION: AVERMITILIS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DR. PETER C. RICHARDSON, PFIZER INC
; STREET: 235 EAST 42ND STREET, 20TH FLOOR
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10017-5755
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,892
; FILING DATE: 6-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SHEYKA, ROBERT F.
; REGISTRATION NUMBER: 31,304
; REFERENCE/DOCKET NUMBER: PC8529B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)573-1189
; TELEFAX: (212)573-1939
; TELEX: N/A
```

```
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 453 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-470-892-4

alignment_scores:
  Quality: 51.00      Length: 42
  Ratio: 1.700       Gaps: 3
Percent Similarity: 71.429 Percent Identity: 33.333

alignment_block:
US-09-209-961-20 x US-08-470-892-4/rev ..
Align seg 1/1 to reverse of: US-08-470-892-4 from: 1 to: 453

17 LeuLeuLeuLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysTh 33
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
432 CTTCGGCTGCTCCAGGATCGCGGC...CACGTCGCCACGACCTT.... 391
: : : : : : : : : : : : : : : : : : : : :
33 rserGlyGlyMetValThrMet.ValLysLysTrpLeuLeuLeuMetThr 49
: : : : : : : : : : : : : : : : : : : : :
390 ....GGACCCAGCTCGCGTCGACGCGGATGTCGAAGCTGACCGCC 345

50 PheMetAlaGlyCysArgGlyMet 57
: : : : : : : : : : : : : : : : : : : : :
344 AGCGTGGTGACCTGTCGGGGCTTG 321

seq_name: /cgn2_5/ptodata/1/ina/5A_COMB.seq:US-08-050-073-2

seq_documentation_block:
; Sequence 2, Application US/08050073
; Patent No. 5567809
; GENERAL INFORMATION:
; APPLICANT: Apple, Raymond J.
; APPLICANT: Begovich, Ann B.
; APPLICANT: Bugawan, Teodorica L.
; APPLICANT: Erlich, Henry A.
; APPLICANT: Griffith, Robert L.
; APPLICANT: Scharf, Stephen J.
; TITLE OF INVENTION: Methods and Reagents for HLA DRBeta DNA
; TITLE OF INVENTION: Typing
; NUMBER OF SEQUENCES: 315
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/050,073
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pety, Douglas A.
; REGISTRATION NUMBER: 35,321
; REFERENCE/DOCKET NUMBER: 8769
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 814-2974
; TELEFAX: (510) 814-2977
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 269 base pairs
```

```
;
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-050-073-2

alignment_scores:
  Quality: 48.50      Length: 56
  Ratio: 1.732      Gaps: 4
  Percent Similarity: 50.000      Percent Identity: 33.929

alignment_block:
US-09-209-961-20 x US-08-050-073-2 ..
Align seg 1/1 to: US-08-050-073-2 from: 1 to: 269

3 ThrSerLeuTrpGlyHisLeuGlyValValVallysAlaAsnGlyLeuLeu11 19
||||| ||||| ||| :||| |||
109 ACAGCGACGTGGGGAGTACCGGGCGGTGACGGAGCTGGGGCGCCTGAT 158
19 eLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrSer.... 34
||| |||||:| :|||
159 G.....CCGAGTACTGGAACAGCCAGAGGACCTCTCTGGA 193
35 .....GlyGlyMetValThrMetValLysLysTrpLeu.LeuLeuMetTh 49
||||| ||||| :||| |||||:|
194 GCAGAGCGGGG.....CCGGGGTGGACACCTATTGCGAGAC 228
49 rPheMetAlaGlyCys 54
| :| :| :| :| :| :|
229 ACAACTAGCGGGCTGT 244

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-050-073-9

seq_documentation_block:
; Sequence 9, Application US/08050073
; Patent No. 5567809
; GENERAL INFORMATION:
; APPLICANT: Apple, Raymond J.
; APPLICANT: Begovich, Ann B.
; APPLICANT: Bugawan, Teodorica L.
; APPLICANT: Erlich, Henry A.
; APPLICANT: Griffith, Robert L.
; APPLICANT: Scharf, Stephen J.
; TITLE OF INVENTION: Methods and Reagents for HLA DRBeta DNA
; TITLE OF INVENTION: Typing
; NUMBER OF SEQUENCES: 315
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/050,073
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Petty, Douglas A.
; REGISTRATION NUMBER: 35,321
; REFERENCE/DOCKET NUMBER: 8769
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 814-2974
; TELEFAX: (510) 814-2977
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
```

```
;
; LENGTH: 269 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-050-073-9

alignment_scores:
  Quality: 48.50      Length: 56
  Ratio: 1.732      Gaps: 4
  Percent Similarity: 50.000      Percent Identity: 33.929

alignment_block:
US-09-209-961-20 x US-08-050-073-9 ..
Align seg 1/1 to: US-08-050-073-9 from: 1 to: 269

3 ThrSerLeuTrpGlyHisLeuGlyValValVallysAlaAsnGlyLeuLeu11 19
||||| ||||| ||| :||| |||
109 ACAGCGACGTGGGGAGTACCGGGCGGTGACGGAGCTGGGGCGCCTGAT 158
19 eLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrSer.... 34
||| |||||:| :|||
159 G.....CCGAGTACTGGAACAGCCAGAGGACCTCTCTGGA 193
35 .....GlyGlyMetValThrMetValLysLysTrpLeu.LeuLeuMetTh 49
||||| ||||| :||| |||||:|
194 GCAGAGCGGGG.....CCGAGGTGGACACCTACTGCGAGAC 228
49 rPheMetAlaGlyCys 54
| :| :| :| :| :| :|
229 ACAACTAGCGGGTGT 244

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-050-073-10

seq_documentation_block:
; Sequence 10, Application US/08050073
; Patent No. 5567809
; GENERAL INFORMATION:
; APPLICANT: Apple, Raymond J.
; APPLICANT: Begovich, Ann B.
; APPLICANT: Bugawan, Teodorica L.
; APPLICANT: Erlich, Henry A.
; APPLICANT: Griffith, Robert L.
; APPLICANT: Scharf, Stephen J.
; TITLE OF INVENTION: Methods and Reagents for HLA DRBeta DNA
; TITLE OF INVENTION: Typing
; NUMBER OF SEQUENCES: 315
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/050,073
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Petty, Douglas A.
; REGISTRATION NUMBER: 35,321
; REFERENCE/DOCKET NUMBER: 8769
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 814-2974
; TELEFAX: (510) 814-2977
; INFORMATION FOR SEQ ID NO: 10:
```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 269 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-050-073-10
```

```
alignment_scores:
  Quality: 48.50      Length: 56
  Ratio: 1.732        Gaps: 4
Percent Similarity: 50.000 Percent Identity: 33.929
```

```
alignment_block:
```

```
US-09-209-961-20 x US-08-050-073-10 ..
```

```
Align seg 1/1 to: US-08-050-073-10 from: 1 to: 269
```

```

3 ThrSerLeuTrpGlyHisLeuGlyValValLysAlaAsnGlyLeuLeu11 19
||||| ||||| ||| :||| |||||
109 ACACGGACGTGGGGAGTACCGGGGTGACGGAGCTGGGGGGCCTGAT 158
: ||||| ||||| ||| :||| |||||
19 eLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrSer.... 34
: ||||| ||||| ||| :||| |||||
159 G.....CCGAGTACTGGAACAGCCAGAGGACCTCTCTGGA 193
35 .....GlyGlyMetValThrMetValLysLysTrpLeu.LeuLeuMetH 49
||||| ||||| ||||| ||||| |||||
194 GCAGAGGCGGG.....CCGCGGTGGACACCTACTGCAGAC 228
49 rPheMetAlaGlyCys 54
| :||| |||||
229 ACAACTACGGGGTTGT 244
```

```
seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-050-073-12
```

```
seq_documentation_block:
; Sequence 12, Application US/08050073
; Patent No. 5567809
; GENERAL INFORMATION:
; APPLICANT: Apple, Raymond J.
; APPLICANT: Begovich, Ann B.
; APPLICANT: Bugawan, Teodorica L.
; APPLICANT: Erlich, Henry A.
; APPLICANT: Griffith, Robert L.
; APPLICANT: Scharf, Stephen J.
; TITLE OF INVENTION: Methods and Reagents for HLA DRbeta DNA
; NUMBER OF SEQUENCES: 315
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/050,073
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Petry, Douglas A.
; REGISTRATION NUMBER: 35,321
; REFERENCE/DOCKET NUMBER: 8769
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 814-2974
; TELEFAX: (510) 814-2977
```

```

; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 269 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-050-073-12
```

```
alignment_scores:
  Quality: 48.50      Length: 56
  Ratio: 1.732        Gaps: 4
Percent Similarity: 50.000 Percent Identity: 33.929
```

```
alignment_block:
```

```
US-09-209-961-20 x US-08-050-073-12 ..
```

```
Align seg 1/1 to: US-08-050-073-12 from: 1 to: 269
```

```

3 ThrSerLeuTrpGlyHisLeuGlyValValLysAlaAsnGlyLeuLeu11 19
||||| ||||| ||| :||| |||||
109 ACACGGACGTGGGGAGTACCGGGGTGACGGAGCTGGGGGGCCTGAT 158
: ||||| ||||| ||| :||| |||||
19 eLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrSer.... 34
: ||||| ||||| ||| :||| |||||
159 G.....CCGAGTACTGGAACAGCCAGAGGACCTCTCTGGA 193
35 .....GlyGlyMetValThrMetValLysLysTrpLeu.LeuLeuMetH 49
||||| ||||| ||||| ||||| |||||
194 GCAGAGGCGGG.....CCGAGGTGGACACCTACTGCAGAC 228
49 rPheMetAlaGlyCys 54
| :||| |||||
229 ACAACTACGGGGTTGT 244
```

```
seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-264-250A-2
```

```
seq_documentation_block:
; Sequence 2, Application US/08264250A
; Patent No. 5663047
; GENERAL INFORMATION:
; APPLICANT: Fumiyu OBATA, et al.
; TITLE OF INVENTION: NOVEL HLA-DR ANTIGEN GENE AND ITS NUCLEOTIDE
; TITLE OF INVENTION: SEQUENCE AND ITS USE
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Office of Sherman and Shalloway
; STREET: 413 N. Washington Street
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM Clone, 8088 Turbo
; OPERATING SYSTEM: MS DOS 5.0
; SOFTWARE: Word Perfect, Version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/264,250A
; FILING DATE: 22-JUN-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/745,444
; FILING DATE: August 15, 1991
; APPLICATION NUMBER: JP A00217209/1990
; FILING DATE: August 20, 1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Richard A. Steinberg
; REGISTRATION NUMBER: 26,588
; REFERENCE/DOCKET NUMBER: HIR-109
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 549-2282
```



ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 60 State Street, suite 510  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/02576  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/205,697  
FILING DATE: 02-Mar-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: BWI-120CPPC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)227-5941  
INFORMATION FOR SEQ ID NO: 27:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 379 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..69  
PCT-US95-02576-27

```

alignment_scores:
  Quality: 49.00      Length: 49
  Ratio: 1.960       Gaps: 1
  Percent Similarity: 51.020  Percent Identity: 28.571

alignment_block:
  US-09-209-961-20 x PCT-US95-02576-27  ..

  Align seg 1/1 to: PCT-US95-02576-27 from: 1 to: 379

      6  TrpGlyHisLeuGlyValValysAlaAsnGlyLeuLeuLeuGlnTh 22
          |||||  |||:|:|:|  :|:|:|:|:|  :|:|
      147  TGGGGATCATGAGCATCTCCCTTACAANAATTAAGCTGTTT.... 191
          |||||  |||  :|:|  :|:|:|:|:|  :|:|

      22  rArgLysProHisThrGlyAsnHisLeuLysThrSerGlyGlyMetValT 39
          ||||  :|:|  :|:|  :|:|:|:|:|  :|:|
      192  .....ACCCACTACCTCACCTCTCTTAAAAACCTCTTTTCAGATTAAAGCT 234
          |||||  :|:|:|:|:|:|  :|:|:|:|:|  :|:|

      39  hrMetValLysLysTrpLeuLeuMetThrPheMetAlaGlyCys 54
          |||||  :|:|:|:|:|  :|:|:|:|:|  :|:|

      235  GRACAGTTACAAGATGCTGGCATCCCTCTCTCTTCTCCCCATATGC 281

seq_name: /cgn2_6/ptodata/1/1na/3A_COMB.seq:US-08-050-073-8

seq_documentation_block:
; Sequence 8, Application US/08050073
; Patent No. 5567809
;
; GENERAL INFORMATION:
;
; APPLICANT: Apple, Raymond J.
; APPLICANT: Begovich, Ann B.
; APPLICANT: Bugawan, Teodorica L.
; APPLICANT: Erlich, Henry A.
; APPLICANT: Griffith, Robert L.
; APPLICANT: Scharf, Stephen J.
;
; TITLE OF INVENTION: Methods and Reagents for HLA DRBeta DNA
; TITLE OF INVENTION: Typing

```

NUMBER OF SEQUENCES: 315  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hoffmann-La Roche Inc.  
STREET: 340 Kingsland Street  
CITY: Nutley  
STATE: New Jersey  
COUNTRY: U.S.A.  
ZIP: 07110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/050,073  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Petry, Douglas A.  
REGISTRATION NUMBER: 35,321  
REFERENCE/DOCKET NUMBER: 8769  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 814-2977  
TELEFAX: (510) 814-2977  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 269 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-050-073-8

alignment\_scores:  
Quality: 47.50 Length: 53  
Ratio: 1.759 Gaps: 3  
Percent Similarity: 50.943 Percent Identity: 33.962  
alignment\_block:  
US-09-209-961-20 x US-08-050-073-8 ..  
Align seg 1/1 to: US-08-050-073-8 from: 1 to: 269  
3 ThrSerLeuTrpGlyHisLeuGlyValValLysAlaAsnGlyLeuLeu1 19  
||||: ||||| ||| :||| |||:  
109 ACACGGACGTGGGGAGTACCGCGGTGACGGAGCTGGGGCGCTGAT 158  
19 eLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrSerGlyG 36  
: ||| |||||: :|||:  
159 G.....CCGAGTACTGGACACCCAGAGGACATCC.... 189  
36 lyMetValThrMetValLysLysTrpLeu.LeuLeuMetThrPheMetAl 52  
||| :|||: |||||: ||| :|||:  
190 ..TGAAGACAGCGCGCGCGGTGGACACCTACTGCAGACACACTACG 237  
52 aglyCys 54  
|||||  
238 GGGTTGT 244

Strd	Orig	Zscore	EScore	Len	Documentation
Sequence					
gb_est16:AA533508	-	67.50	159.16	322	! AA533508 nk81b11.s1 NCI_CGAP_SC
gb_est16:AA531068	-	67.50	158.52	343	! AA531068 nk74g07.s1 NCI_CGAP_SC
gb_est19:AA780257	+	62.50	149.58	241	! AA780257 af5521.s1 Soares test
gb_est13:AA346195	+	62.50	148.85	259	! AA346195 ES525496 Fetal heart
gb_est16:AA559041	-	64.50	148.28	447	! AA559041 n11le01.s1 NCI_CGAP_Br
gb_est13:AV059058	-	62.50	147.88	286	! AV059058 AV059058 Mus musculus
gb_est18:AA723992	-	59.00	145.52	152	! AA723992 ai03a12.s1 Soares para
gb_gss3:AA090205	+	62.50	143.66	277	! AA090205 HS_3009.AL F06_T7 CIT
gb_est25:AI283571	+	60.50	143.10	278	! AI283571 qh5d06.x1 Soares_NFL
gb_est34:AV139690	-	60.50	142.88	284	! AV139690 AV139690 Mus musculus
gb_est16:NH8961	-	58.00	142.55	159	! NH8961 K6909F Human fetal heart
gb_gss5:AO183363	-	61.50	142.04	394	! AO183363 HS_3142.B2 B10.MR CIT
gb_gss4:AO173665	+	62.50	141.91	510	! AO173665 HS_3204.B1.D04.T7 CIT
gb_gss4:AO177242	+	62.50	141.60	526	! AO177242 HS_3033.B1.E05.MR CIT
gb_gss9:AO626156	+	63.00	141.38	373	! AO626156 CITBI-EL-2657Y22.TR CIT
gb_est28:AI499767	-	60.00	141.38	291	! AI499767 tm92c09.x1 NCI_CGAP_Br
gb_est10:AA137792	-	61.00	140.48	406	! AA137792 zk92q06.x1 Soares pres
gb_gss3:AO162637	-	63.50	140.45	752	! AO162637 mgxb0013D12r CUGI Rice
gb_est25:AI317195	+	62.00	140.17	535	! AI317195 ui35f12.y1 Soares mous
gb_est24:AI236687	+	60.00	140.06	331	! AI236687 ES2733249 Normalized r
gb_gss7:AO134690	+	61.50	140.04	479	! AO134690 RPTC11-103H8.T7 RPTC1
gb_est34:AV120743	-	59.50	139.85	299	! AV120743 AV120743 Mus musculus
gb_gss9:AO628604	+	61.00	139.68	439	! AO628604 CITBI-EL-2656M14.TR CIT
gb_est19:AA743308	-	60.00	138.12	400	! AA743308 nv20c08.s1 NCI_CGAP_GC
gb_est36:AV198363	-	59.50	137.95	367	! AV198363 AV198363 Yuji Kohara U
gb_gss3:AO09147	+	59.50	137.86	363	! AO09147 HS_3047.AL.B08.MF CIT
gb_gss8:AO575593	-	63.00	137.81	861	! AO575593 nbxb0087H17r CUGI Rice
gb_est5:NA0843	-	59.50	137.47	377	! NA0843 yw82h11.r1 Soares placed
gb_gss9:AO618671	+	60.50	137.47	482	! AO618671 HS_S169.A2.D03.T7A.RPC
gb_est17:AA651891	+	60.00	137.31	433	! AA651891 ns39a11.s1 NCI_CGAP_GC
gb_est37:AI972966	+	58.50	137.20	303	! AI972966 wr45h06.x1 NCI_CGAP_Pt
gb_gss4:AO173685	+	60.50	137.12	499	! AO173685 HS_3204.B1.F03.T7 CIT
gb_gss11:AO817096	+	59.00	136.84	355	! AO817096 HS_5553.AL.B11.SP6.RH
gb_est32:AI736225	+	57.50	136.62	251	! AI736225 ea41d02.y1 Elmerita S5-
gb_est32:AI757353	+	57.50	136.54	253	! AI757353 ea32a12.y1 Elmerita S5-
gb_est24:AI184605	+	60.00	136.39	474	! AI184605 qf46b08.x1 Soares test
gb_gss6:AO405011	+	60.50	136.36	71	! AO405011 HS_5048.B2.G08.T7 RPTC
gb_gss9:AO623632	-	60.50	136.33	539	! AO623632 CITBI-EL-265518.TF CIT
gb_est32:AV059456	-	57.50	136.26	260	! AV059456 AV059456 Mus musculus
gb_gss5:AO274684	+	61.50	136.16	700	! AO274684 RPTC-5-113314Sp6 RPTC-
gb_est20:AA824297	+	60.00	136.13	486	! AA824297 aj30a02.s1 Soares test
gb_est30:AA03851	-	58.00	136.08	299	! AA03851 AV003851 Mus musculus

```

22 rArgLysProHistr.....GlyAsnHisLeuL 32
   ||||| |||||
173 ACAAAGGACCACTCTGATGCACTTGAAGCAGCATTTGCTTATAGA 124

32 ysthrSerGlyMetValThrMetValLysLysTrpLeuLeuMet 48
   ||||| |||||
123 AAAACATATCTGGAGTAAAGAACAGAGTAAGAACAGTCTTCTGATG 74

49 ThrPheMet 51
   |||||
73 GAGTTCATG 65

```

seq\_name: gb\_est16:AA551068

```

seq_documentation_block: 343 bp mRNA EST 09-SEP-1997
LOCUS AA551068
DEFINITION nk74907.s1 NCI-CGAP_Sch1 Homo sapiens cDNA clone IMAGE:1019292 3'
similar to TR:G1039472 G1039472 PHEROMONE RECEPTOR VN7.1; mRNA
sequence.
ACCESSION AA551068
NID 92321320
VERSION AA551068.1 GI:2321320
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 343)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On May 5, 1995 this sequence version replaced gi:797888.

```

Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: Stratagene, Inc., David B. Krizman,  
 Ph.D.  
 CDNA Library Arraying: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 www.bio.llnl.gov/dbdp/image.html

Insert Length: 886 Std Error: 0.00  
 Seq primer: -40ml3 fwd. Et from Amersham  
 High quality sequence stop: 276.

```

FEATURES
  source
    1..343
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /map="y"
      /clone="IMAGE:1019292"
      /clone_lib="NCI-CGAP_Sch1"
      /tissue_type="Schwannoma tumor"
      /lab_host="SOLR (kanamycin resistant)"
      /note="Vector: Bluescript SK-; Site 1: EcoRI; Site 2:
XhoI: Cloned unidirectionally. Primer: Oligo dt. Two
pooled bulk Schwannoma tumors. 5' adaptor sequence: 5'
GAATTCGGACGAG 3' 3' adaptor sequence: 5'
CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.2 kb."

```

```

BASE COUNT 83 a 93 c 61 g 106 t
ORIGIN

alignment_scores:
  Quality: 67.50 Length: 53
  Ratio: 1.985 Gaps: 1
Percent Similarity: 64.151 Percent Identity: 30.189

```

alignment\_block:  
 US-09-209-961-20 x AA551068/rev ..

Align seg 1/1 to reverse of: AA551068 from: 1 to: 343

```

6 TrpGlyHisLeuGlyValValLysAlaAsnGlyLeuLeuLeuGlnTh 22
   ||||| ||||| ||||| |||||
223 TGGGAGAGGCGTGTGCTGTAAGATATCAGAAATTTCTTCTGTGCTGA 174

22 rArgLysProHistr.....GlyAsnHisLeuL 32
   ||||| |||||
173 ACAAAGGACCACTCTGATGCACTTGAAGCAGCATTTGCTTATAGA 124

32 ysthrSerGlyMetValThrMetValLysLysTrpLeuLeuMet 48
   ||||| |||||
123 AAAACATATCTGGAGTAAAGAACAGAGTAAGAACAGTCTTCTGATG 74

49 ThrPheMet 51
   |||||
73 GAGTTCATG 65

```

seq\_name: gb\_est19:AA780257

```

seq_documentation_block: 241 bp mRNA EST 05-FEB-1998
LOCUS AA780257
DEFINITION af55c11.s1 Soares total fetus Nb2HF8.9w Homo sapiens cDNA clone
IMAGE:1035572 3' similar to gb:X06617 40S RIBOSOMAL PROTEIN S11
(HUMAN); mRNA sequence.
ACCESSION AA780257
NID 92839588
VERSION AA780257.1 GI:2839588
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 241)
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Krimman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,
Martin, J., Moore, B., Scheinberg, K., Steptoe, M., Tan, F.,
Theising, B., White, Y., Wyllie, T., Waterston, R. and Wilson, R.
WashU-NCI human EST Project
Unpublished (1997)
COMMENT On Jan 9, 1998 this sequence version replaced gi:785632.

```

Contact: Wilton RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Trace considered overall poor quality  
 Seq primer: -40ml3 fwd. Et from Amersham  
 High quality sequence stop: 1.

```

FEATURES
  source
    1..241
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="IMAGE:1035572"
      /clone_lib="Soares_total_fetus_Nb2HF8.9w"
      /dev_stage="8-9 weeks"
      /lab_host="DH10B"
      /note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from pooled 8-9 week
(total) fetus material with a Not I - oligo(dT) primer [5'
TGTACCAATCTGAAGTGGAGCGCCCTTAATTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was

```



breast tumor tissue, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7n3 vector. This library is the normalized version of NCI-CGAP-Brl.1. Library was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 145 a 85 c 79 g 138 t  
ORIGIN

alignment\_scores:  
Quality: 64.50 Length: 58  
Ratio: 1.897 Gaps: 2  
Percent Similarity: 58.621 Percent Identity: 29.310

alignment\_block:  
US-09-209-961-20 x AA559041/rev ..  
Align seg 1/1 to reverse of: AA559041 from: 1 to: 447

5 LeuTrpGlyHisLeuGlyValValLysAlaAsnGlyLeuLeu..... 18

436 GTTTGGAACACACAGAGAGAGTTGTGTCATGGTGTATCTGTGAGC 387

19 .....IleLeuGlnThrArgLysProHisThr..... 27

386 AACCTCCTATGTCTCTGTGATGACCTGCTACTCATCTCCTGTGTAAAT 337

28 .....GlyAsnHisLeuLysThrSerGlyGlyMetValThr 39

336 CAAGAATAAGACCTGGGAAGTTGTGTGAAGTCTCCCTATTAAATGATTCT 287

40 MetValLysLysTrpLeuLeuLeu 47

286 CACGTGAGAAGATGGATCTTACTG 263

seq\_name: gb\_est32:AV059058

seq\_documentation\_block:  
LOCUS AV059058 286 bp mRNA EST 23-JUN-1999  
DEFINITION AV059058 Mus musculus pancreas C57BL/6J adult Mus musculus cDNA clone 1810057B02, mRNA sequence.

ACCESSION AV059058

NID 95158805

VERSION AV059058.1 GI:5158805

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 286)

Akahiru, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T., Hara, A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M., Kawai, J., Kikuchi, N., Kojima, Y., Matsuyama, T., Niitsuma, H., Oda, H., Owa, C., Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Tateo, M., Tomaru, Y., Tomimaga, N., Watanabe, S., Yagame, M., Yamamura, T., Yokota, T., Yoshino, M., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.

RIKEN Mouse ESTs

Unpublished (1999)

On Jun 22, 1998 this sequence version replaced gi:3247139.

CONTACT: Chie Owa

Genome Science Laboratory

RIKEN

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

Tel: 81-298-36-9145

Fax: 81-298-36-9098

Email: genome-res@rtc.riken.go.jp

Thermotabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA

(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))  
Transcriptional sequencing: A method for DNA sequencing using RNA polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))  
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES  
source

1. .286  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/map="xp11.2"  
/clone="1810057B02"  
/clone\_lib="Mus musculus pancreas C57BL/6J adult"  
/sex="male"  
/tissue\_type="pancreas"  
/dev\_stage="adult"  
69 a 56 c 59 g 102 t

BASE COUNT  
ORIGIN

alignment\_scores:  
Quality: 62.50 Length: 31  
Ratio: 2.976 Gaps: 2  
Percent Similarity: 67.742 Percent Identity: 48.387

alignment\_block:  
US-09-209-961-20 x AV059058/rev ..

Align seg 1/1 to reverse of: AV059058 from: 1 to: 286

26 HistThrGlyAsnHisLeuLysThrSerGlyGlyMetValThrMetValThr 42

139 CACAGTGGACATGGTCTCAGGACTCTCTAGGT.....CTTCTGAA 99

42 sLysTrpLeuLeuMetThrPheMetAlaGlyCysArgGly 56

98 GAAGTGG.....CCAGGGGATTCAGGACGCTGTGAGGG 62

seq\_name: gb\_est18:AA723992

seq\_documentation\_block:  
LOCUS AA723992 152 bp mRNA EST 31-DEC-1998  
DEFINITION a103a12.s1 Soares\_parathyroid\_tumor\_NbHPA Homo sapiens cDNA clone 1341694 3', mRNA sequence.

ACCESSION AA723992

NID 92741699

VERSION AA723992.1 GI:2741699

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 152)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT On Sep 12, 1996 this sequence version replaced gi:1401009.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert\_Strausberg@nih.gov

CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

[www-bio.llnl.gov/dbp/image/image.html](http://www-bio.llnl.gov/dbp/image/image.html)

Insert Length: 1061 Std Error: 0.00

Seq primer: -40ml3 fwd. ET from Amersham.

FEATURES  
Location/Qualifiers

```

1. 152
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="1341694"
/clone_lib="Soares.parathyroid_tumor_NbHPA"
/tissue_type="parathyroid tumor"
/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: parathyroid gland; Vector: pT7T3D
(Pharmacia) with a modified polylinker; Site:1;
Site:2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer
(5'-
TGTTACCACTCTGAAGTGGAGCGCGCCACCAATTTTTTTTTTTTTTTTTTTT
T-3')], double-stranded cDNA was size selected, ligated to
Eco RI adapters (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of a modified pT7T3
vector (Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. RNA from sporadic parathyroid
adenomas was kindly provided by Dr. Stephen Marx, National
Institute of Diabetes and Digestive and Kidney Diseases,
NIH."
BASE COUNT      37 a      42 c      36 g      37 t
ORIGIN

alignment_scores:
    Quality:      59.00      Length:      34
    Ratio:        2.458      Gaps:      1
Percent Similarity: 70.588      Percent Identity: 38.235

alignment_block:
US-09-209-961-20 x AA723992/rev ..

Align seg 1/1 to reverse of: AA723992 from: 1 to: 152

6 TrpGlyHisLeuGlyValVallysAlaAsnGlyLeuLeu..... 18
   ||| |||||||:||||:||||:||||| |||
148 TGGCACCATCTGGCGCTGTGTCAGGCCCATGGAGGTGGAGGCCACCA 99

19 .lIeLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrSerg 35
   |||||||:||||:||||:||||| ||||| :|||
98 CATCCTTAAGCCATCAGTAGTATAGTGGTCCACCCTGCATGTGAAGG 49

35 ly 35
   ||
48 GG 47

seg_name: gb_gss3:AQ090205

seg_documentation_block:
LOCUS AQ090205 430 bp DNA GSS 25-AUG-1998
DEFINITION HS_3009_AL_F06_T7 CIT Approved Human Genomic Sperm Library D Homo
          sapiens genomic clone Plate=3009 Col=11 Row=K, genomic survey
          sequence.
ACCESSION AQ090205
NID 93459116
VERSION AQ090205.1 GI:3459116
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
          Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 430)
Mahairas,G.G., Zackrone,K.D., Smith,T., Tipton,S., Schmidt,S.,
Traicoff,R., Abajian,C., Blanchard,A., West,A. and Hood,L.P.
Construction of a Characterized Clone Resource for Genomic
Sequencing: Generation and Preliminary Analysis of 20,000 Sequence
Tagged Connectors
Unpublished (1997)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Contact: Mahairas GG, Wallace JC, Hood L

```





JOURNAL  
COMMENT

Unpublished (1996)  
On Apr 14, 1993 this sequence version replaced gi:638496.

Contact: Liew CC  
Department of Laboratory Medicine and Pathobiology  
University of Toronto  
Banting Institute, 100 College St., Toronto, Ontario, M5G1L5  
Tel: 4169788758  
Fax: 4169785650  
Email: liewcc@utcc.utoronto.ca  
Seq primer: GAATTAACCCCTCACTAAAGGG.

## FEATURES

source  
1. .159  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="K6909"  
/clone\_lib="Human fetal heart, Lambda ZAP Express"  
/lab\_host="E. coli XL1-Blue"  
/note="Vector: Lambda ZAP Express; Site\_1: EcoRI; Site\_2:  
XhoI; mRNA was purified from human fetal hearts (8-10  
weeks). cDNA was synthesized using a XhoI-Oligo dT  
adaptor-primer. EcoRI adaptors were ligated, followed by  
digestion with XhoI, for directional cloning into  
predigested lambda ZAP Express."

BASE COUNT 57 a 35 c 56 g 11 t  
ORIGIN

## alignment\_scores:

Quality: 58.00 Length: 35  
Ratio: 2.762 Gaps: 2  
Percent Similarity: 60.000 Percent Identity: 42.857

## alignment\_block:

US-09-209-961-20 x N88961/rev ..

Align seg 1/1 to reverse of: N88961 from: 1 to: 159

24 LysProHisThrGlyAsnHisLeuLysrSerGlyGlyMetValThrMe 40

||||| : : : : : ||||| : : : : :  
100 AAGCCTTATCTCCCTGCATCCTT.....GGCTGCTTCTTGG 60

40 tValLysTrpLeuLeuMetThrPheMetAlaGly.....C 54

: : : : : ||||| ||||| |||||  
59 GCTGTTTCAGTGGCTCTCTTCGCCACCTTCGCCGCCGACATGGCGCCT 10

54 ysArg 55  
|||||

9 GCCGC 5

seq\_name: gb\_gss5:AQ183363

seq\_documentation\_block:

LOCUS AQ183363 394 bp DNA GSS 31-OCT-1998  
DEFINITION HS\_3142\_B2\_B10\_MR CIT Approved Human Genomic Sperm Library D Homo  
sapiens genomic clone Plate=3142 Col=20 Row=D, genomic survey  
sequence.

ACCESSION AQ183363

NID G3580730

VERSION AQ183363.1 GI:3580730

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 394)

MAHAIAS,G.G., Wallace,J.C., Smith,K., Swartzell,S., Furlong,J.,

Shaker,R., Schmidt,S., Traicoff,R. and Hood,L.E.

Construction of a Characterized Clone Resource for Genomic

Sequencing

Unpublished (1998)

JOURNAL

COMMENT

Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887

Email: jwallace@u.washington.edu  
Sequence Tagged Connector  
Plate: 3142 row: D column: 20  
Class: BAC ends

High quality sequence stop: 394.

## FEATURES

Location/Qualifiers  
1. .394  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Plate=3142 Col=20 Row=D"  
/clone\_lib="CIT Approved Human Genomic Sperm Library D"  
/sex="male"  
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in  
E-Coli DH10B"

BASE COUNT 117 a 64 c 67 g 146 t  
ORIGIN

## alignment\_scores:

Quality: 61.50 Length: 45  
Ratio: 2.050 Gaps: 2  
Percent Similarity: 66.667 Percent Identity: 37.778

## alignment\_block:

US-09-209-961-20 x AQ183363/rev ..

Align seg 1/1 to reverse of: AQ183363 from: 1 to: 394

18 LeuLeuGlnThrArgLysProHisThrGlyAsnHisLeu...LysTh 33

||||| : : : : : ||||| : : : : :  
153 CTTTITTTTAAACAATGAGTGGCAATGACGGCTAACCATATCAAGAAAC 104

33 rSerGlyGlyMetValThrMetVal.....LysLysTrpLeuLeuLeu 48

: : : : : ||||| ||||| |||||  
103 TCAATGCAGATATATCACTATGTGTAGAAAAAATAGCTTCTACTGC 54

48 etThrPheMetAlaGlyCysArgGlyMetIleTyr 59

: : : : : ||||| : : : : :  
53 TCACAGCTCTTCATATGTGTGTCATGCTCTTTTATAC 19

seq\_name: gb\_gss4:AQ173665

seq\_documentation\_block:

LOCUS AQ173665 510 bp DNA GSS 16-OCT-1998  
DEFINITION HS\_3204\_B1\_D04\_T7 CIT Approved Human Genomic Sperm Library D Homo  
sapiens genomic clone Plate=3204 Col=7 Row=H, genomic survey  
sequence.

ACCESSION AQ173665

NID G3571032

VERSION AQ173665.1 GI:3571032

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 510)

MAHAIAS,G.G., Wallace,J.C., Smith,K., Swartzell,S., Furlong,J.,

Shaker,R., Schmidt,S., Traicoff,R. and Hood,L.E.

Construction of a Characterized Clone Resource for Genomic

Sequencing

Unpublished (1998)

JOURNAL

COMMENT

Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu  
 Sequence Tagged Connector  
 Plate: 3204 row: H column: 7  
 Class: BAC ends  
 High quality sequence stop: 510.

## FEATURES

source  
 1. .510  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="plate=3204 Col=7 Row=H"  
 /clone\_lib="CIT Approved Human Genomic Sperm Library D"  
 /sex="male"  
 /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 198 a 85 c 111 g 116 t  
 ORIGIN

alignment\_scores:  
 Quality: 62.50 Length: 52  
 Ratio: 2.016 Gaps: 2  
 Percent Similarity: 59.615 Percent Identity: 32.692

alignment\_block:  
 US-09-209-961-20 x AQ173665 ..

Align seg 1/1 to: AQ173665 from: 1 to: 510

6 TrpGlyHisLeuGlyValValLysAlaAsnGlyLeuLeuLeuGlnTh 22  
 |||||  
 235 TGGAAATCATCGC.....ATAGAATCGAATGGAATATCATCGAATGC 278

22 rArgLysProHisThrGlyAsnHisLeuLysThrSerGlyGlyMetValt 39  
 |||||  
 279 TCGAATCGAATCAACATCAACGGAATCAACGGAATATGGAATGGAAT 328

39 hrMetValLysLysTrpLeuLeuMetThrPheMetAlaGlyCysArg 55  
 ::|||  
 329 CG.....AAGAGAATCATCGAATGGAATCGAATGGAATCATCTAAT 369

56 GlyMet 57  
 |||||

370 GGAATG 375

seq\_name: gb\_gss4:AQ177242

seq\_documentation\_block:  
 LOCUS AQ177242 526 bp DNA GSS 16-OCT-1998  
 DEFINITION HS\_3033\_B1\_E05\_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate-3033 Col=9 Row=J, genomic survey sequence.

ACCESSION AQ177242  
 NID 93574609  
 VERSION AQ177242.1 GI:3574609  
 KEYWORDS GSS.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
 Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 526)

AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Furlong,J., Shaker,R., Schmidt,S., Traicoff,R. and Hood,L.E.  
 TITLE Construction of a Characterized Clone Resource for Genomic Sequencing  
 JOURNAL Unpublished (1998)

COMMENT  
 Contact: Mahairas GG, Wallace JC, Hood L  
 High Throughput Sequencing Center  
 University of Washington  
 401 Queen Anne Avenue North, Seattle, WA 98109, USA  
 Tel: (206) 616-3618  
 Fax: (206) 616-3887  
 Email: jwallace@u.washington.edu

Sequence Tagged Connector  
 Plate: 3033 row: J column: 9  
 Class: BAC ends  
 High quality sequence stop: 526.  
 Location/Qualifiers  
 1. .526  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="plate=3033 Col=9 Row=J"  
 /clone\_lib="CIT Approved Human Genomic Sperm Library D"  
 /sex="male"  
 /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 198 a 93 c 118 g 115 t 2 others  
 ORIGIN

alignment\_scores:  
 Quality: 62.50 Length: 52  
 Ratio: 2.016 Gaps: 2  
 Percent Similarity: 59.615 Percent Identity: 32.692

alignment\_block:  
 US-09-209-961-20 x AQ177242 ..

Align seg 1/1 to: AQ177242 from: 1 to: 526

6 TrpGlyHisLeuGlyValValLysAlaAsnGlyLeuLeuLeuGlnTh 22  
 |||||  
 264 TGGAAATCATCGC.....ATAGAATCGAATGGAATATCATCGAATGC 307

22 rArgLysProHisThrGlyAsnHisLeuLysThrSerGlyGlyMetValt 39  
 |||||  
 308 TCGAATCGAATCAACATCAACGGAATCAACGGAATATGGAATGGAAT 357

39 hrMetValLysLysTrpLeuLeuMetThrPheMetAlaGlyCysArg 55  
 ::|||  
 358 CG.....AAGAGAATCATCGAATGGAATCGAATGGAATCATCTAAT 398

56 GlyMet 57  
 |||||

399 GGAATG 404

seq\_name: gb\_gss9:AQ626156

seq\_documentation\_block:  
 LOCUS AQ626156 607 bp DNA GSS 16-JUN-1999  
 DEFINITION CITBI-EI-2657P22.TR CITBI-EI Homo sapiens genomic clone 2657P22, genomic survey sequence.

ACCESSION AQ626156  
 NID 95088548  
 VERSION AQ626156.1 GI:5088548  
 KEYWORDS GSS.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 607)

AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and Venter,J.C.  
 TITLE Use of BAC End Sequences from CalTech Libraries for Sequence-Ready Map Building  
 JOURNAL Unpublished (1997)

COMMENT  
 Other\_GSSs: CITBI-EI-2657P22.TF  
 Contact: Shaying Zhao, William Nierman, Mark Adams  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: hbeetigr.org  
 Clones are available from Research Genetics (info@resgen.com). BAC end search page:

http://www.tigr.org/tdb/hungen/bac\_end\_search/bac\_end\_search.html.  
Seq primer: M13 Reverse  
Class: BAC ends.

FEATURES  
    source  
        location/Qualifiers  
            1..607  
            /organism="Homo sapiens"  
            /db\_xref="taxon:9606"  
            /clone="2657P22"  
            /clone\_lib="CITBI-E1"  
            /sex="male"  
            /cell\_type="sperm"  
            /note="Vector: pBelosAC11; Site\_1: EcoRI; Site\_2: EcoRI;  
            CalTech Human BAC Library D"  
BASE COUNT    226 a    66 c    114 g    201 t  
ORIGIN

alignment\_scores:  
    Quality:    63.00        Length:    30  
    Ratio:      3.000        Gaps:      0  
    Percent Similarity: 70.000    Percent Identity: 40.000

alignment\_block:  
US-09-209-961-20 x AQ626156 ..

Align seg 1/1 to: AQ626156 from: 1 to: 607  
30 HisLeuLysThrSerGlyGlyMetValThrMetValLysLysTrpLeuLe 46  
    |||||:          |||||:          :  :  |||  ||  
70 CATTTAATGAATGGGGGAGTAGTAAGCTGTTATAGTATTGGAAACT 119  
46 uLeuMetThrPheMetAlaGlyCysArgGlyMetIleTyr 59  
    |||||:|||||  |||||:          :  :  |||  ||  
120 ATGGGTAACATTTAACACAGTGTGTTAACAGTTTATCTAT 159

OM of: US-09-209-961-21 to: GenEmbl:\* out\_format : pfs  
Date: Dec 27, 1999 2:13 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-1998 CompuGen Ltd.

Command line parameters:  
-MODEL=frame+p2n\_model -DEV=xlp  
-Q/cgn2\_1/USPFO\_spool/US09209961/runat\_21211999\_101043\_17010/app\_query.fasta.1  
-DB=GenEmbl -QWMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500  
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -LOCALIGN=200 -THR\_SCORE=score -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pfs -NORM=stat -USER=US09209961 -NCPU=6 -ICPU=3 -NO\_XLPXY  
-WAIT -THREADS=1

## Search information block:

Query: US-09-209-961-21  
Query length: 53  
Database: GenEmbl:\*  
Database sequences: 780561  
Database length: 2137953050  
Search time (sec): 1994.760000

## score\_list:

Sequence	Strd	Orig	Zscore	EScore	Len	Documentation
gb_vi:AF134203	+	273.00	681.76	5.7e-30	436	AF134203 Porcine circovirus PCV
gb_vi:AF135393	+	273.00	681.56	5.7e-30	438	AF135393 Porcine circovirus PCV
gb_vi:AF118097	+	273.00	677.59	9.7e-29	700	AF118097 Porcine circovirus ty
gb_vi:AF085695	+	273.00	669.57	2.7e-29	1768	AF085695 Porcine circovirus ty
gb_vi:AF086835	+	273.00	669.57	2.7e-29	1768	AF086835 Porcine circovirus st
gb_vi:AF086835	+	273.00	669.57	2.7e-29	1768	AF086835 Porcine circovirus st
gb_vi:AF086835	+	273.00	669.57	2.7e-29	1768	AF086835 Porcine circovirus st
gb_vi:AF086835	+	273.00	669.57	2.7e-29	1768	AF086835 Porcine circovirus st
gb_vi:AF109397	+	273.00	669.57	2.7e-29	1768	AF109397 Bovine circovirus, co
gb_vi:AF112862	+	273.00	669.57	2.7e-29	1768	AF112862 Porcine circovirus ty
gb_vi:AF055391	+	270.00	661.98	7.2e-29	1768	AF055391 Porcine circovirus ty
gb_vi:PCAJ3185	+	269.00	659.45	9.9e-29	1768	AF223185 Porcine circovirus DN
gb_vi:AF109399	+	267.00	654.39	1.9e-28	1768	AF109399 Porcine circovirus ty
gb_vi:AF027217	+	266.00	651.86	2.6e-28	1768	AF027217 Porcine circovirus st
gb_vi:AF147751	+	265.00	649.33	3.6e-28	1768	AF147751 Porcine circovirus ty
gb_vi:AF154679	+	262.00	641.74	9.6e-28	1768	AF154679 Porcine circovirus, c
gb_vi:AF166528	+	259.00	634.15	2.5e-27	1768	AF166528 Porcine circovirus, c
gb_vi:AF109398	+	249.00	608.85	6.5e-26	1768	AF109398 Porcine circovirus ty
gb_vi:AF117753	+	240.00	608.85	6.5e-26	1768	AF117753 Porcine circovirus ty
gb_vi:AF055393	+	210.00	510.18	2.0e-20	1767	AF055393 Porcine circovirus ty
gb_vi:AF055394	+	210.00	510.18	2.0e-20	1767	AF055394 Porcine circovirus ty
gb_vi:AF071879	+	72.50	162.34	0.4846	1759	AF071879 Porcine circovirus, c
gb_vi:AF012107	+	70.00	156.01	1.09	1759	AF012107 Porcine circovirus, c
gb_vi:PCCOMGEN	+	68.00	150.95	2.09	1759	Y09921 Porcine circovirus comp
gb_vi:PC049186	+	68.00	150.95	2.09	1759	Y09186 Porcine circovirus, com
gb_pr1:HUMSATRA	+	70.00	145.14	4.40	2970	L01057 Human (clone FRI-6) sat
gb_pr1:D50001S06	+	70.00	144.34	4.87	6760	D50006 Human DNA for alpha-pla
gb_in2:AF079059	-	58.00	139.20	9.42	368	AF079059 Mesobuthus martensii c
gb_sts:HS320XB5	+	57.50	137.59	11.59	383	Z51365 H.sapiens (D383516) DNA
gb_in2:AR091022	+	57.00	136.79	12.84	363	AF091022 E. coli phage phiX174
gb_pl1:YSCITR1	+	64.00	136.34	13.60	2953	D90352 S.cerevisiae IFR1 gene
gb_pl2:AF004707	+	57.50	136.25	13.76	447	AF004707 Gigaspora albida strain
gb_vi:AF106408	-	57.00	136.10	14.02	393	AF106408 HIV-1 isolate T2B5041
gb_pr2:HSU85H7	-	72.50	135.93	14.33	37054	I 273900 Human DNA sequence frc
gb_pr1:AF012770	-	62.00	135.01	16.12	1919	AF012770 Homo sapiens gene for
gb_vi:AF106409	-	56.00	133.64	19.23	390	AF106409 HIV-1 isolate T2B5043
gb_vi:AF106398	-	56.00	133.57	19.40	393	AF106398 HIV-1 isolate T2B5030
gb_vi:AF106452	-	56.00	133.57	19.40	393	AF106452 HIV-1 isolate T2B5092
gb_cm:CFDOPRX09	+	55.50	133.48	19.61	343	X64970 C.familialis mRNA DOPCRX
gb_ro:RAVINKL	-	59.00	132.38	22.60	1083	M11884 Rat low molecular weight
gb_ro:AF017085	-	63.50	132.25	22.98	4091	AF017085 Mus musculus BAP-135
gb_in2:AF034556	+	55.50	131.75	24.37	417	AF034556 Osccharinella rochebru

gb\_vi:HV228318 - 54.50 131.03 26.87 340 ! AJ228318 Human immunodeficie  
gb\_vi:AF106390 - 55.00 130.97 27.06 396 ! AF106390 HIV-1 isolate T2B50  
gb\_vi:AF106444 - 55.00 130.91 27.29 399 ! AF106444 HIV-1 isolate T2B50

seq\_name: gb\_vi:AF134203

seq\_documentation\_block:  
LOCUS AF134203 436 bp ss-DNA VRL 02-JUN-1999

DEFINITION Porcine circovirus PCV-2A amplified sequence.

ACCESSION AF134203

NID 94959580

VERSION AF134203.1 GI:4959580

KEYWORDS porcine circovirus.

SOURCE porcine circovirus.

ORGANISM Viruses; ssDNA viruses; Circoviridae; Circovirus.

REFERENCE 1 (bases 1 to 436)

A PCR assay for detecting and characterizing type-2 circovirus in

Hamel, A.L., Lin, L., Sachvie, C., Grudeski, E. and Nayar, G.P.S.

plgs, cattle, bison and sheep with various clinical syndromes

Unpublished

REFERENCE 2 (bases 1 to 436)

Hamel, A.L., Lin, L., Sachvie, C., Grudeski, E. and Nayar, G.P.S.

Direct Submission

Submitted (10-MAR-1999) Virology Laboratory, Manitoba Agriculture,

Veterinary Services, 545 University Crescent, Winnipeg, Manitoba

R3T 5S6, Canada

FEATURES

Location/Qualifiers

1..436

/organism="porcine circovirus"

/strain="PCV-2A"

/db\_xref="taxon:46221"

/note="isolated from sheep (ewe)"

1..436

misc\_feature

/note="PCR amplification product using oligonucleotide

primers derived from porcine circovirus from pigs

deposited in Genbank Accession Number AF027217"

BASE COUNT 107 a 83 c 103 g 143 t

ORIGIN

alignment\_scores:

Quality: 273.00 Length: 53

Ratio: 5.151 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-209-961-21 x AF134203 ..

Align seg 1/1 to: AF134203 from: 1 to: 436

1 MetValPheillelleHisLeuGlyPheLysTrpGlyValPheLysIlely 17

155 ATGGTTTATTTATTCATTAGTGGTTTAAAGTGGGGCTTTAAGATTAA 234

17 sPheSerGluLeuTyrlleHisGlyTyrrThrAspIleValValIleSer 34

235 ATTCCTCAATTGTACATACATAGTTTACACGGATATTAGTCTGCTGCG 284

34 alPheThrValPheGluArgSerAlaGluAlaTyrrValValHisIleSer 50

285 TATTTACTGTTTTCGACGACAGTCCGAGGCGCTACGTGTCACCATTTCT 334

51 ArgGlyLeu 53

335 AGAGGTTTG 343

seq\_name: gb\_vi:AF135393

seq\_documentation\_block:

LOCUS AF135393 438 bp ss-DNA VRL 02-JUN-1999

DEFINITION Porcine circovirus PCV-2A amplified sequence.

ACCESSION AF135393

```

NID          94959706
VERSION      AF135393.1  GI:4959706
KEYWORDS     porcine circovirus.
SOURCE       porcine circovirus.
ORGANISM     Viruses; ssDNA viruses; Circoviridae; Circovirus.
REFERENCE    1 (bases 1 to 438)
AUTHORS      Hamel,A.L., Lin,L., Sachvie,C., Grudeski,E. and Nayar,G.P.S.
TITLE        A PCR assay for detecting and characterizing type-2 circovirus in
JOURNAL      pigs, cattle, bison and sheep with various clinical syndromes
              Unpublished
REFERENCE    2 (bases 1 to 438)
AUTHORS      Hamel,A.L., Lin,L., Sachvie,C., Grudeski,E. and Nayar,G.P.S.
TITLE        Direct Submission
JOURNAL      Submitted (17-MAR-1999) Virology Laboratory, Veterinary Services,
              Manitoba Agriculture, 545 University Crescent, Winnipeg, Manitoba
              R3T 5S6, Canada
FEATURES     source
              1..438
              /organism="porcine circovirus"
              /strain="PCV-2A"
              /db_xref="taxon:46221"
              /note="isolated from bison"
              1..438
              /note="PCR amplification product using oligonucleotide
              primers derived from porcine circovirus from pig deposited
              in GenBank Accession Number AF027217"
BASE COUNT   106 a 83 c 104 g 145 t
ORIGIN
alignment_scores
  Quality: 273.00      Length: 53
  Ratio: 5.151        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000
alignment_block:
  US-09-209-961-21 x AF135393  ..
  Align seg 1/1 to: AF135393 from: 1 to: 438
      1 MetValPheIleIleHisLeuGlyPheLysTrpGlyValPheLysIleLy 17
      186 ATGGTGTATTATTATTAGGTTAAAGTGGGGGCTTTAAGATTAA 235
      17 sPheSerGluLeuTyriIleHisGlyTyThrAspIleValValLeuValV 34
      236 ATTCTCTGAATTCATACATACATGTTACACGGATATTGTAGTCTGTCG 285
      34 alPheThrValPheGluArgSerAlaGluAlaTyrValValHisIleSer 50
      286 TATTTACTGTTTTCGACGACGAGTGGCGAGGCTACGTGGTCCACATTCT 335
      51 ArgGlyLeu 53
      336 AGAGGTTTG 344
  seq_name: gb_vi:AF118097
seq_documentation_block:
  LOCUS      AF118097      700 bp      DNA      VRL      27-JUL-1999
  DEFINITION Porcine circovirus type 2 isolate IAF-4370 unknown gene.
  ACCESSION  AF118097
  NID        95596427
  VERSION    AF118097.1  GI:5596427
  KEYWORDS   porcine circovirus type 2.
  SOURCE     porcine circovirus type 2
  ORGANISM   Viruses; ssDNA viruses; Circoviridae; Circovirus.
  REFERENCE 1 (bases 1 to 700)
  AUTHORS    Ouardani,M., Wilson,L., Jette,R., Montpetit,C. and Dea,S.
  TITLE      Multiplex PCR for detection and typing of porcine circoviruses
  JOURNAL    J. Clin. Microbiol. (1999) In press
REFERENCE    2 (bases 1 to 700)
AUTHORS      Ouardani,M., Wilson,L., Jette,R., Montpetit,C. and Dea,S.
TITLE        Direct Submission
JOURNAL      Submitted (06-JAN-1999) CRMB, INRS-IAF, 531, Boulevard des
              Prairies, Laval, Que H7N 4Z3, Canada
FEATURES     source
              1..700
              /organism="porcine circovirus type 2"
              /isolate="IAF-4370"
              /db_xref="taxon:85708"
              /note="virus isolated from lung, lymph node, spleen and
              tonsil tissue from pigs affected by postweaning
              multisystemic wasting syndrome"
              complement(42..>700)
              /note="ORF2"
              /codon_start=3
              /product="unknown"
              /protein_id="AAD45581.1"
              /db_xref="PID:95596428"
              /db_xref="GI:5596428"
              /translation="BSHLGQILRRRLVPHRRVRRKNGIENLRISFTGCTVKA
              TTVTPSWAYDMRENIDDFVPGGTNKISIPFEYRKRKXAVFWPSPVQGBRG
              VGSTAVIDDDNFVTATATIDPYVINYSSRHTIPQPFHSRIFTKRPVLDSTIDYFQ
              PNKNRQLMRLQTSRNVDHVGIGTAFENSKYDQDYNIRVTMYVQFREFNLKDPPLKP
              "
BASE COUNT   156 a 127 c 210 g 207 t
ORIGIN
alignment_scores
  Quality: 273.00      Length: 53
  Ratio: 5.151        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000
alignment_block:
  US-09-209-961-21 x AF118097  ..
  Align seg 1/1 to: AF118097 from: 1 to: 700
      1 MetValPheIleIleHisLeuGlyPheLysTrpGlyValPheLysIleLy 17
      24 ATGGTGTATTATTATTAGGTTAAAGTGGGGGCTTTAAGATTAA 73
      17 sPheSerGluLeuTyriIleHisGlyTyThrAspIleValValLeuValV 34
      74 ATTCTCTGAATTCATACATACATGTTACACGGATATTGTAGTCTGTCG 123
      34 alPheThrValPheGluArgSerAlaGluAlaTyrValValHisIleSer 50
      124 TATTTACTGTTTTCGACGACGAGTGGCGAGGCTACGTGGTCCACATTCT 173
      51 ArgGlyLeu 53
      174 AGAGGTTTG 182
  seq_name: gb_vi:AF055392
seq_documentation_block:
  LOCUS      AF055392      1768 bp      DNA      circular      VRL      13-SEP-1998
  DEFINITION Porcine circovirus Type II from Canada, complete genome.
  ACCESSION  AF055392
  NID        93598808
  VERSION    AF055392.1  GI:3598808
  KEYWORDS   porcine circovirus.
  SOURCE     porcine circovirus
  ORGANISM   Viruses; ssDNA viruses; Circoviridae; Circovirus.
  REFERENCE 1 (bases 1 to 1768)
  AUTHORS    Meehan,B.M., McNeilly,F., Todd,D., Kennedy,S., Jewhurst,V.A.,
              Ellis,J.A., Hassard,L.E., Clark,E.G., Haines,D.M. and Allan,G.M.
  TITLE      Characterization of novel circovirus DNAs associated with wasting
              syndromes in pigs
  JOURNAL    J. Gen. Virol. 79 (Pt 9), 2171-2179 (1998)

```

```

MEDLINE 98418498
REFERENCE 2 (bases 1 to 1768)
AUTHORS Meehan,B.M.
TITLE Direct Submission
JOURNAL Submitted (26-MAR-1998) Brian M Meehan, Veterinary Science, The Queen's University of, Belfast, Stoney Road, Belfast, Northern Ireland, BT4 3SD, U.K.
FEATURES
    source
        1..1768
            /organism="porcine circovirus"
            /specific_host="Sus scrofa"
            /strain="Porcine circovirus Type II"
            /db_xref="taxon:46221"
            /clone="Imp.1010-Stoon ECORI No.14"
            /country="Canada"
            complement(101..190)
            /codon_start=1
            /product="ORF6"
            /protein_id="AAC35319.1"
            /db_xref="PID:g3598819"
            /db_xref="GI:3598819"
            /translation="MASSTPASPAPSDILSSVPQSORPPGRWT"
            103..210
            /codon_start=1
            /product="ORF10"
            /protein_id="AAC35317.1"
            /db_xref="PID:g3598817"
            /db_xref="GI:3598817"
            /translation="MSTAQEGVVTVVRLTVYKVRERRVLKMPFFLLOR"
            261..320
            /codon_start=1
            /product="ORF7"
            /protein_id="AAC35318.1"
            /db_xref="PID:g3598818"
            /db_xref="GI:3598818"
            /translation="MAAGAVSSAVTPPWIRHS"
            complement(311..439)
            /note="predicted 4.6 kDa protein"
            /codon_start=1
            /product="ORF9"
            /protein_id="AAC35315.1"
            /db_xref="PID:g3598815"
            /db_xref="GI:3598815"
            /translation="MWLGSSILLAGHVAAEVLPCCRCRSALVILTAHFFRFQL"
            398..1342
            /codon_start=1
            /product="putative Rep protein (ORF1)"
            /protein_id="AAC35309.1"
            /db_xref="PID:g3598809"
            /db_xref="GI:3598809"
            /translation="MPSKNGRSGPQPHKRWFTLNPNSEDEKKIRELPISLFDVFI
            VGEENEGRTPHLQGFANFKQTFNKVYLGARCHEIKAKGTQONKEYCKEKN
            LIECGAPRSQGSLSYAVSTLLESGSLTVAEQHPVTFVRFGLAELLKVSQGM
            OKRDWNTVHVIVGPPCGCKSKWAANFADPTTYWKPNNKWDVGHGEVVDVDFY
            GHLPWDLRLCLDRYPLTVETKGTVPFLARSILITSNQPLEWYSTAVPAVEALYR
            RITSLVFNKATLQSTEEGGQFVTLSPPCPEFFPEINY"
            674..679
            complement(704..1018)
            /note="predicted 11.9 kDa protein"
            /codon_start=1
            /product="ORF3"
            /protein_id="AAC35311.1"
            /db_xref="PID:g3598811"
            /db_xref="GI:3598811"
            /translation="MTIPLVSRWFPVCGFRVCKISSPFAFTTPRPHNDVIGLPI
            TLLHFAHQKFSQPAEISDKRYRVLICNGHQIPALQQGTHSSRQVTPLSRSRSTF
            NK"
            complement(733..912)
            /note="predicted 6.5 kDa protein"
            /codon_start=1
            /product="ORF4"
            /protein_id="AAC35312.1"
            /db_xref="PID:g3598812"

polyA_signal
CDS
    1 MetValPheIleIleHisLeuGlyPheLysTrpGlyValPheLysIleIle 17
    1363 ATGGTATTATTATTATTCATTTAGGTGTTAAGTGGGGGTCTTTAAGATTAA 1412
    17 sPheSerGluLeuTyrIleHisGlyTyrThrAspIleValValLeuValV 34
    1413 ATTCCTGAATGTACATACATGTTTACACGGATATTAGTCCTGCTGTCG 1462
    34 aPheThrValPheGluArgSerAlaGluAlaTyrValValHisIleSer 50
    1463 TATTACTGTTTTCGAACGACGCGCGAGCCCTACGTGGTCCACATTCC 1512
    51 ArgGlyLeu 53
    1513 AGAGGTTTG 1521
    seq_name: gb_v1:AF085695

BASE COUNT 451 a 361 c 495 g 461 t
ORIGIN

alignment_scores
    Quality: 273.00 Length: 53
    Ratio: 5.151 Gaps: 0
    Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block
    US-09-209-961-21 x AF055392 ..
    Align seg 1/1 to: AF055392 from: 1 to: 1768
    1 MetValPheIleIleHisLeuGlyPheLysTrpGlyValPheLysIleIle 17
    1363 ATGGTATTATTATTATTCATTTAGGTGTTAAGTGGGGGTCTTTAAGATTAA 1412
    17 sPheSerGluLeuTyrIleHisGlyTyrThrAspIleValValLeuValV 34
    1413 ATTCCTGAATGTACATACATGTTTACACGGATATTAGTCCTGCTGTCG 1462
    34 aPheThrValPheGluArgSerAlaGluAlaTyrValValHisIleSer 50
    1463 TATTACTGTTTTCGAACGACGCGCGAGCCCTACGTGGTCCACATTCC 1512
    51 ArgGlyLeu 53
    1513 AGAGGTTTG 1521
    seq_name: gb_v1:AF085695

```

```
seq_documentation_block:
LOCUS AF085695 1768 bp DNA circular VRL 30-SEP-1998
DEFINITION Porcine circovirus strain 412, complete genome.
ACCESSION AF085695
NID 93668362
VERSION AF085695.1 GI:3668362
KEYWORDS
SOURCE
ORGANISM
porcine circovirus.
VIRUSES: ssDNA viruses; Circoviridae; Circovirus.
REFERENCE
1 (bases 1 to 1768)
AUTHORS Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.
TITLE Emergence of a new porcine circovirus.
JOURNAL Unpublished
2 (bases 1 to 1768)
AUTHORS Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.
TITLE Direct Submission
JOURNAL Submitted (21-AUG-1998) Veterinary Infectious Disease Organization,
University of Saskatchewan, 120 Veterinary Road, Saskatoon,
Saskatchewan S7N 5E3, Canada
FEATURES
source
location/Qualifiers
1..1768
/organism="porcine circovirus"
/strain="412"
/db_xref="taxon:46221"
51..995
/codon_start=1
/product="p35.8"
/protein_id="AAC61860.1"
/db_xref="PID:3668363"
/db_xref="GI:3668363"
/translation="WPSKNGBSGPQPHKRWVFTLNPSDEKIKIRLPISLFDYFI
VGEENEGRPHLOGFANFYKQFENKVKWILGARCHEKAKGDDQNKCYCSREGN
LLIECGAPQQRSDLSSTAVSLIESGILTVAKOHPVTFVKNRGLAELLKVSCKM
OKRDKTNVHTFVGPCCGSKWAFANPETTYKPPKNKWDYHGKVVVIDDFY
GWLWDDLLRLCDRYPLAVTKTGTVPLFARSILITSNQPLEWYSSTAVPAVEALYR
RTSLVFWKNATKQSTEGGQFVTLSPCCPEPPEYIN"
complement(357..671)
/codon_start=1
/product="p12.1"
/protein_id="AAC61865.1"
/db_xref="PID:3668368"
/db_xref="GI:3668368"
/translation="WVTIPPLVFRNFPVCGFRVCKISSPFAETTPRWPHNEVYIGFPI
TLHFPFAHFQKFSQAEIFDKRIRVLLCNHONPALQOQTHSRQVTPLSLRSSTF
NK"
complement(386..565)
/codon_start=1
/product="p6.6"
/protein_id="AAC61866.1"
/db_xref="PID:3668369"
/db_xref="GI:3668369"
/translation="MKCTLVFQSRFCIFPLTFKSSASPRKFLTNVTGCCFAIVTRIPL
SNKVLTAVDRLRCP"
553..732
/codon_start=1
/product="p6.6"
/protein_id="AAC61861.1"
/db_xref="PID:3668364"
/db_xref="GI:3668364"
/translation="WYTSLWGHVGVKANGLLILQTRPHTGNHLKTSQGMVTKKKW
LLMTFMAGCRNII"
833..970
/codon_start=1
/product="p5.0"
/protein_id="AAC61862.1"
/db_xref="PID:3668365"
/db_xref="GI:3668365"
/translation="MYLLNCPSCRSLSSEDFYLGILECYKTIHGGRGPFVRHPFPPM
P"
1016..1177
/codon_start=1
```

```
/product="p6.2"
/protein_id="AAC61863.1"
/db_xref="PID:3668366"
/db_xref="GI:3668366"
/translation="WVFIHLGFKWGVFKIKFSELYIHGYTDIVLVVFTVFFERSAEA
YVYHSRGL"
complement(1034..1735)
/codon_start=1
/product="p27.9"
/protein_id="AAC61864.1"
/db_xref="PID:3668367"
/db_xref="GI:3668367"
/translation="MTYIPRRYRRRRPRSHLQILRRRPWLVPHPHRYWRKNGI
FNTLRSTFGYTVKRTVTPSWAVDMRFKIDFVPPGGTGNKISIPFYIRKVK
VEFWPCSPITQDGRGVSTAVILDDNFVKATALTVDYPVNVYSRRHTIPOPFSVHSRY
FTFPKVLDTIDYFQPNKRNQLWLRLQISGNDVHVLGTAFENSKYDQDYNIRVTHY
VQFREFNLADPPLEP"
BASE COUNT 463 a 362 c 481 g 462 t
ORIGIN

alignment_scores:
Quality: 273.00 Length: 53
Ratio: 5.151 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-209-961-21 x AF085695 ..
Align seg 1/1 to: AF085695 from: 1 to: 1768
1 MetValPheIleIleHisLeuGlyPheLysTrpGlyValPheLysIlely 17
1016 ATGTTTATATATCATTTAGGTTCAAGTGGGGGCTTTAGATTAA 1065
17 sPheSerGluLeuTyrlleHisGlyTyThrAspileValValLeuValv 34
1066 ATTCTCTGAATTGACATACATGTTACACGGATATTGTAGTCCTGGTCG 1115
34 alpheThrValPheGluArgSerAlaGluAlaTyrrValValHisIleSer 50
1116 TATTACTGTTTTCGACGACGAGTCCGAGGCTACGTGTCACATTTC 1165
51 ArgGlyLeu 53
1166 AGAGTTTG 1174
seq_name: gb_vi:AF086834
seq_documentation_block:
LOCUS AF086834 1768 bp DNA circular VRL 29-SEP-1998
DEFINITION Porcine circovirus strain B9, complete genome.
ACCESSION AF086834
NID 93661515
VERSION AF086834.1 GI:3661515
KEYWORDS
SOURCE
ORGANISM
porcine circovirus.
VIRUSES: ssDNA viruses; Circoviridae; Circovirus.
REFERENCE
1 (bases 1 to 1768)
AUTHORS Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.
TITLE Emergence of a new porcine circovirus
JOURNAL Unpublished
2 (bases 1 to 1768)
AUTHORS Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.
TITLE Direct Submission
JOURNAL Submitted (26-AUG-1998) VIDO, University of Saskatchewan, 120
Veterinary Road, Saskatoon, Saskatchewan S7N 5E3, Canada
FEATURES
source
location/Qualifiers
1..1768
/organism="porcine circovirus"
/strain="B9"
/db_xref="taxon:46221"
```

```
CDS          51..995
/codon_start=1
/product="P35.8"
/protein_id="AAC61737.1"
/db_xref="PID:g3661516"
/db_xref="GI:3661516"
/translation="MPSKNGRSGPQPHKRWFTLNPPSEDERKKIRELPISLFDYFI
VGEENEGRTPHLOGFANFVKOTENKVKWLGARCHIEKAKGTDOONKEVCSEKGN
LLECGAPRSQORSLSLSTAVSLLESGLVTAKQHPVTFVKNFGLAELLKVSQGM
QRDKTNVHFIVGPGCGSKWAANFANPETTYWPKPKNWDGKHGKVVVDDFY
RITSLVFWKNTQSTEEGGQFVTLSPPCPEFFEYNY"
/complement(1034..1735)
/codon_start=1
/codon_start=1
/product="P27.9"
/protein_id="AAC61738.1"
/db_xref="PID:g3661517"
/db_xref="GI:3661517"
/translation="MTYPRRRRRRRRSHRSHLQILRRRPWLVPVPHRYRWRKNGI
ENTLSRTFGYTVKRTVTPSWAVDMRENIIDFVPPGGGTNKISIPFEYRIRKVK
VEFWPCSPITQDGRGVSTAVILDDNFVTKATALTVDYPVYNSRHTIPQPFYSHSRY
FTPKPVDSTIDIFQPNKRNQLWLRLQTSRNVHDVHGLGTAFENSKYIDQDYNIRVTM
VQFREFNLKDPPLKP"
BASE COUNT   463 a   357 c   482 g   466 t
ORIGIN

alignment_scores:
  Quality: 273.00      Length: 53
  Ratio: 5.151        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-209-961-21 x AF086834 ..
Align seg 1/1 to: AF086834 from: 1 to: 1768
1 MetValPheIleIleHisLeuGlyPheLysTrpGlyValPheLysIlely 17
1016 ATGGTATTTTATTATTCATTAGGTTTAAAGTGGGGGCTTTAAAGATTAA 1065

17 sPheSerGluLeuTyrlleHisGlyTyThrAspIleValValLeuValv 34
1066 ATTCCTGAATTGACATACATAGTTACACGGATATTGTAGTCCGGTCG 1115

34 alPheThrValPheGluArgSerAlaGluAlaTyValValHisIleSer 50
1116 TATTTACTGTTTTTCGACGACGAGCTAGCTAGCTAGCTAGCTAGCTAGCT 1165

51 ArgGlyLeu 53
1166 AGAGGTTTG 1174

seq_name: gb_vi:AF086835

seq_documentation_block:
LOCUS      AF086835      1768 bp      DNA      circular      VRL
DEFINITION Porcine circovirus strain 9741, complete genome.
ACCESSION  AF086835
NID        93661518
VERSION    AF086835.1 GI:3661518
KEYWORDS   porcine circovirus.
SOURCE     porcine circovirus
ORGANISM   porcine circovirus
REFERENCE  1 (bases 1 to 1768)
AUTHORS    Wang,L., Willson,P., Chow,B., Gibbons,E. and Babiuk,L.
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 1768)
AUTHORS    Wang,L., Willson,P., Chow,B., Gibbons,E. and Babiuk,L.
TITLE      Direct Submission
JOURNAL    Submitted (26-AUG-1998) VIDO, University of Saskatchewan, 120

FEATURES             Location/Qualifiers
     source            1..1768
                        /organism="porcine circovirus"
                        /strain="9741"
                        /db_xref="taxon:46221"
                        51..995
     CDS               /codon_start=1
                        /product="P35.8"
                        /protein_id="AAC61739.1"
                        /db_xref="PID:g3661519"
                        /db_xref="GI:3661519"
                        /translation="MPSKNGRSGPQPHKRWFTLNPPSENERKKIRKIRELPISLFDYFI
                        VGEENEGRTPHLOGFANFVKOTENKVKWLGARCHIEKAKGTDOONKEVCSEKGN
                        LLECGAPRSQORSLSLSTAVSLLESGLVTAKQHPVTFVKNFGLAELLKVSQGM
                        QRDKTNVHFIVGPGCGSKWAANFANPETTYWPKPKNWDGKHGKVVVDDFY
                        RITSLVFWKNTQSTEEGGQFVTLSPPCPEFFEYNY"
                        complement(1034..1735)
                        /codon_start=1
                        /product="P27.9"
                        /protein_id="AAC61740.1"
                        /db_xref="PID:g3661520"
                        /db_xref="GI:3661520"
                        /translation="MTYPRRRRRRRRSHRSHLQILRRRPWLVPVPHRYRWRKNGI
                        ENTLSRTFGYTVKRTVTPSWAVDMRENIIDFVPPGGGTNKISIPFEYRIRKVK
                        VEFWPCSPITQDGRGVSTAVILDDNFVTKATALTVDYPVYNSRHTIPQPFYSHSRY
                        FTPKPVDSTIDIFQPNKRNQLWLRLQTSRNVHDVHGLGTAFENSKYIDQDYNIRVTM
                        VQFREFNLKDPPLKP"
     BASE COUNT       464 a   360 c   480 g   464 t
     ORIGIN

alignment_scores:
  Quality: 273.00      Length: 53
  Ratio: 5.151        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-209-961-21 x AF086835 ..
Align seg 1/1 to: AF086835 from: 1 to: 1768
1 MetValPheIleIleHisLeuGlyPheLysTrpGlyValPheLysIlely 17
1016 ATGGTATTTTATTATTCATTAGGTTTAAAGTGGGGGCTTTAAAGATTAA 1065

17 sPheSerGluLeuTyrlleHisGlyTyThrAspIleValValLeuValv 34
1066 ATTCCTGAATTGACATACATAGTTACACGGATATTGTAGTCCGGTCG 1115

34 alPheThrValPheGluArgSerAlaGluAlaTyValValHisIleSer 50
1116 TATTTACTGTTTTTCGACGACGAGCTAGCTAGCTAGCTAGCTAGCTAGCT 1165

51 ArgGlyLeu 53
1166 AGAGGTTTG 1174

seq_name: gb_vi:AF086835

seq_documentation_block:
LOCUS      AF086835      1768 bp      DNA      circular      VRL
DEFINITION Porcine circovirus strain 9741, complete genome.
ACCESSION  AF086835
NID        93661518
VERSION    AF086835.1 GI:3661518
KEYWORDS   porcine circovirus.
SOURCE     porcine circovirus
ORGANISM   porcine circovirus
REFERENCE  1 (bases 1 to 1768)
AUTHORS    Wang,L., Willson,P., Chow,B., Gibbons,E. and Babiuk,L.
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 1768)
AUTHORS    Wang,L., Willson,P., Chow,B., Gibbons,E. and Babiuk,L.
TITLE      Direct Submission
JOURNAL    Submitted (26-AUG-1998) VIDO, University of Saskatchewan, 120
```



TITLE Emergence of a new porcine circovirus  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1768)  
AUTHORS Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.  
TITLE Direct Submission  
JOURNAL Submitted (26-AUG-1998) VIDO, University of Saskatchewan, 120  
Veterinary Road, Saskatoon, Saskatchewan S7N 5E3, Canada  
FEATURES  
Source  
1. 1768  
/organism="porcine circovirus"  
/strain="W226"  
/db\_xref="taxon:46221"  
51. 995  
/product="p35.8"  
/protein\_id="AAC61741.1"  
/db\_xref="PID:93661522"  
/db\_xref="GI:3661522"  
/translation="MPSKNGRSGPQPHRRWFTLNNPSEDERKKIRELPISLFDYFI  
VGEENEGRTPLHGFANFVKOTFNKVLFGARCHIEKAKGTDOONKEYCSKEG  
LLIEGAPRSQGSQSDLSFVSTLLESGILTVAKOHPVTFVKNRGLAEKLVSKGM  
QKRDKNTHVHIVGPPCGCKSKWAANFANPETTYWKPKNKWDGYSYHGEVYVDDFY  
GWLPPDLDRLCDRPLTVTKGGTVPLFARSILITSNQTPLEWYSSTAVPAVEADYR  
RITSLVFNKNATKQSTEEGGQFVTLSPPCPEFFEYNI"  
complement(1034..1735)  
/codon\_start=1  
/product="p27.9"  
/protein\_id="AAC61742.1"  
/db\_xref="PID:93661523"  
/db\_xref="GI:3661523"  
/translation="MTYPRRRYRRRRRSHLQILRRRPLVHPRHRYRWRKNGKI  
FNRLSRFGYVTKRTVTPSWAVDMREKIDDFVPGGNTKISIPETRYIRKVK  
VEFPCSPITQDGRGVSTAVILDONFVKATALYDPVNYSSRHTIPOPYSHSRY  
ETPKVLSTIDYDFPNKRNQLRLQTSNGNVHVLGTAFENSKIDQDINIRVTMI  
VQREFNKLDPLEP"  
BASE COUNT 463 a 363 c 480 g 462 t  
ORIGIN

alignment\_scores:  
Quality: 273.00 Length: 53  
Ratio: 5.151 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-209-961-21 x AF086836 ..  
Align seg 1/1 to: AF086836 from: 1 to: 1768  
1 MetValPheIleIleHisLeuGlyPheIleYsrpGlyValPheLysIleLy 17  
|||||  
1016 ATGCTTTTATTATTCATTAGGTTCAAGTGGGGGCTTTAAGATTAA 1065  
|||||  
17 sPheSerGluLeuTyriIleHisGlyTyThrAspIleValValLeuValV 34  
|||||  
1066 ATTCTCTGAATTGACATACATGTTACACGGATATTGATCTCTGGTCG 1115  
|||||  
34 alPheThrValPheGluArgSerAlaGluAlaTyThrValValHisIleSer 50  
|||||  
1116 TATTACTGTTTCGACGAGTCCGAGGCTACGTGGTCCACATTTC 1165  
|||||  
51 ArgGlyLeu 53  
|||||  
1166 AGAGGTTTG 1174  
|||||

seq\_name: gb\_vi:AF109397

seq\_documentation\_block:  
LOCUS AF109397 1768 bp DNA circular VRL 02-FEB-1999  
DEFINITION Bovine circovirus, complete genome.  
ACCESSION AF109397  
NID 94210313  
VERSION AF109397.1 GI:4210313

KEYWORDS  
SOURCE bovine circovirus.  
ORGANISM bovine circovirus  
REFERENCE Viruses; ssDNA viruses; Circoviridae; Circovirus.  
AUTHORS 1 (bases 1 to 1768)  
TITLE Hamel, A.L. and Nayar, G.P.S.  
Nucleotide sequence of a circovirus detected in cattle with various  
clinical syndromes  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1768)  
AUTHORS Hamel, A.L. and Nayar, G.P.S.  
TITLE Direct Submission  
JOURNAL Submitted (27-NOV-1998) Virology Laboratory, Veterinary Diagnostic  
Laboratory, Manitoba Agriculture, 545 University Crescent,  
Winnipeg, Manitoba R3T 5S6, Canada  
REFERENCE 3 (bases 1 to 1768)  
AUTHORS Hamel, A.L. and Nayar, G.P.S.  
TITLE Direct Submission  
JOURNAL Submitted (02-FEB-1999) Virology Laboratory, Veterinary Diagnostic  
Laboratory, Manitoba Agriculture, 545 University Crescent,  
Winnipeg, Manitoba R3T 5S6, Canada  
REMARK Sequence update by submitter  
COMMENT On Feb 2, 1999 this sequence version replaced gi:4028609.  
FEATURES  
Source  
1. 1768  
/organism="bovine circovirus"  
/viroion  
/db\_xref="taxon:85542"  
/note="BCV  
sequence was obtained from several overlapping PCRs using  
DNA extracted from various tissues from cattle; similar to  
Porcine circovirus sequence presented in GenBank Accession  
Number AF027217."  
13. 35  
/repeat\_region  
/rpt\_type=tandem  
/rpt\_unit=13. 18  
51. 995  
/note="ORF-1: similar to Porcine circovirus ORF1 encoded  
by the sequence presented in GenBank Accession Number  
AF027217."  
/codon\_start=1  
/product="putative Rep and coat protein"  
/protein\_id="AAD11928.1"  
/db\_xref="PID:94210314"  
/db\_xref="GI:4210314"  
/translation="MPSKNGRSGPQPHRRWFTLNNPSEDERKKIRELPISLFDYFI  
VGEENEGRTPLHGFANFVKOTFNKVLFGARCHIEKAKGTDOONKEYCSKEG  
LLIEGAPRSQGSQSDLSFVSTLLESGILTVAKOHPVTFVKNRGLAEKLVSKGM  
QKRDKNTHVHIVGPPCGCKSKWAANFADPETTYWKPKNKWDGYSYHGEVYVDDFY  
GWLPPDLDRLCDRPLTVTKGGTVPLFARSILITSNQTPLEWYSSTAVPAVEALYR  
RITSLVFNKNATQSTEEGGQFVTLSPPCPEFFEYNI"  
complement(357..671)  
/note="similar to Porcine circovirus ORF-3 encoded by the  
sequence presented in GenBank Accession Number AF027217."  
/codon\_start=1  
/product="ORF-3"  
/protein\_id="AAD11930.1"  
/db\_xref="PID:94210315"  
/db\_xref="GI:4210315"  
/translation="MTYIPLVSRWFVPCGFRVCKISSPFAFTTPRPHNDVYICLPI  
TLHPFAHFOKFSQPAEISDKRYRLVLCNHQHPALQGGTHSSRQVPLSLRSRSTF  
NK"  
complement(386..565)  
/note="similar to Porcine circovirus ORF-4 encoded by the  
sequence presented in GenBank Accession Number AF027217."  
/codon\_start=1  
/product="ORF-4"  
/protein\_id="AAD11931.1"  
/db\_xref="PID:94210316"  
/db\_xref="GI:4210316"  
/translation="MTCTLVFQSRFCIFPLTFKSSASPRKFLTNVTCSSATVTRLP  
SNKLVTAVDRLRCP"

polyA\_signal  
CDS

CDS

CDS

```

CDS
553..732
/codon_start=1
/product="ORF-12"
/protein_id="AAD11927.1"
/db_xref="PID:94028610"
/db_xref="GI:4028610"
/translation="MYTSLWGLGVVYKANGLLILOTRKPHGTGNHLETSGGVTMVKKK
LLLMTFMAGCRGMIV"
complement(688..753)
/codon_start=1
/product="ORF-8"
/protein_id="AAD11935.1"
/db_xref="PID:94028618"
/db_xref="GI:4028618"
/translation="WDIDHTSVSDHPTAASHKSHQ"
983..988
complement(989..1033)
/codon_start=1
/product="ORF-11"
/protein_id="AAD11937.1"
/db_xref="PID:94028620"
/db_xref="GI:4028620"
/translation="MNNKNHYEVIKTKQ"
1016..1177
/codon_start=1
/product="ORF-5"
/protein_id="AAD11932.1"
/db_xref="PID:94028615"
/db_xref="GI:4028615"
/translation="VVFIIHLGKVGKFKFSELYIHGYTDIVLVWVTFERSAEA
YVYHISRLG"
complement(1022..1027)
complement(1256..1735)
/notes="similar to Porcine circovirus ORF-2 encoded by the
sequence presented in GenBank Accession Number AF027217."
/codon_start=1
/product="ORF-2"
/protein_id="AAD11929.1"
/db_xref="PID:94028612"
/db_xref="GI:4028612"
/translation="MTYPRRRYRRRTPRSHLQQLRRPMLVHPRHRYRWRKNGI
FNRLSRFGYVATVTPSWAVDMRMENIDDFVPGGYNKISIPFYRIRKVK
VFEPSPITQDGRGVGTAVILDNFVTKATALTYDPYVNTSSRTIQQPSYHSR"
complement(1522..1611)
/codon_start=1
/product="ORF-6"
/protein_id="AAD11933.1"
/db_xref="PID:94028616"
/db_xref="GI:4028616"
/translation="MASSTPASPAPSDILSRLPQSERPPGRWT"
1524..1631
/codon_start=1
/product="ORF-10"
/protein_id="AAD11936.1"
/db_xref="PID:94028619"
/db_xref="GI:4028619"
/translation="MSTAQEGVLTVVALTVPKVRERRVLRKMPFFLLQR"
1682..1741
/codon_start=1
/product="ORF-7"
/protein_id="AAD11934.1"
/db_xref="PID:94028617"
/db_xref="GI:4028617"
/translation="MAAGAGPSSAVTPPPWIRHS"
join(1750..1768,1..13)
join(1762..1768,1..2)
/notes="putative; similar to the nonanucleotide motif of
Porcine circovirus presented in GenBank Accession Numbers
AF027217 and U49186"
452 a 362 c 493 g 461 t
BASE COUNT
ORIGIN

```

```

alignment_scores:
  Quality: 273.00      Length: 53
  Ratio: 5.151        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-209-961-21 x AF109397 ..

Align seg 1/1 to: AF109397 from: 1 to: 1768

1 MetValPheLleIleHisLeuGlyPheLysTrpGlyValPheLysIlely 17
1016 ATGGTTTATTATTATTCATTAGGTTTAAGTGGGGGCTTTTAAGATTAA 1065
17 sPheSerGluLeuTyIleHisGlyTyThrAspIleValValLeuValV 34
1066 APTCTCTGAATTGTATACATACATGTTACACGGATATTAGTCTGTCG 1115
34 alphaThrValPheGluArgSerAlaGluAlaTyValValHisIleSer 50
1116 TATTACGTGTTTCGACGACGAGTCCGAGGCGCTACGTGTCACATTCT 1165
51 ArgGlyLeu 53
1166 AGAGGTTTG 1174

seq_name: gb_vi:AF112862

seq_documentation_block:
LOCUS AF112862 1768 bp DNA circular VRL 06-JAN-1999
DEFINITION Porcine circovirus type 2-B, complete genome.
ACCESSION AF112862
NID 94106915
VERSION AF112862.1 GI:4106915
KEYWORDS
SOURCE
ORGANISM
porcine circovirus type 2-B.
porcine circovirus type 2-B.
Viruses; ssDNA viruses; Circoviridae; Circovirus;
REFERENCE
1 (bases 1 to 1768)
AUTHORS
Hamel,A.L. and Nayar,G.P.S.
TITLE
Nucleotide sequence of four different isolates of porcine
circovirus detected in pigs with various clinical syndromes
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 1768)
AUTHORS
Hamel,A.L. and Nayar,G.P.S.
TITLE
Direct Submission
JOURNAL
Submitted (10-DEC-1998) Virology Laboratory, Manitoba Agriculture,
545 University Crescent, Winnipeg, Manitoba R3T 5S6, Canada
FEATURES
source
1..1768
/organism="porcine circovirus type 2-B"
/db_xref="taxon:85709"
/notes="sequence obtained from several overlapping PCRs
using DNA extracted from lung, mesenteric lymph node and
tonsil of pig; similar to Porcine circovirus sequence
presented in GenBank Accession Number AF027217; type-B
designation is based upon restriction endonuclease
digestion pattern"
13..35
/rept_type=tandem
/rept_unit=13..18
51..995
/notes="ORF-1"
/codon_start=1
/product="putative Rep and coat protein"
/protein_id="AAD03086.1"
/db_xref="PID:94106918"
/db_xref="GI:4106918"
/translation="MPSRKNGSGPQPKRWVFTLNNSSEDEKKIRELPISLFDYFI
VGEENEGRTPHLQGFANFKVKTENKVKILGARCHIEKAKGTDOONKEYCSKEGN
LLIECGAPSGQORSLSLSTAVSLLESGLVVAEQHPVTFYRFRGLAELLKYSKGM
QKRWKNTNHHVIVGPPCGKSKWAANFADPTTYKPRKRWGYSGEVVVDDFY
GWLPPDLLRLCDRIPLVTETKGTVPFLFARSILITSNQTPLEWYSSTAVPAVEALYR

```





```
ORGANISM      porcine circovirus
REFERENCE     Viruses: ssDNA viruses; Circoviridae; Circovirus.
AUTHORS      1 (bases 1 to 1768)
TITLE        Morozov, I.
JOURNAL      Direct Submission
             Submitted (02-JAN-1998) Morozov I., Veterinary Medical Research
             Institute, Iowa State University, 1802 Elwood Dr. VMRI Ames, Iowa
             50011, U.S.A.
REFERENCE     2 (bases 1 to 1768)
AUTHORS      Morozov, I., Sirdinamitr, T., Sorden, S.D., Halbur, P.G., Morgan, M.K.,
             Yoon, K.J. and Paul, P.S.
TITLE        Detection of a novel strain of Porcine circovirus in pigs with
             Postweaning Multisystemic Wasting Syndrome (PMWS)
JOURNAL      Unpublished
FEATURES     Location/Qualifiers
             source
             1..1768
             /organism="porcine circovirus"
             /virion
             /isolate="ISU-31"
             /specific host="Sus scrofa"
             /db_xref="taxon:46221"
             51..995
             /note="ORF1"
             /codon_start=1
             /protein_id="CA11157.1"
             /db_xref="PID:e1310034"
             /db_xref="PID:g3293041"
             /db_xref="GI:3293041"
             /translation="MPSKNGRSGPQPKRWFTLNNPSEDERKKIRELPISLFDYFI
             VGEENEGRTPHLOGFANFKVKTQFNKRWLGARCHIEKAKGTQQNKYCSKEGN
             LLIECGAPRSQGRSDLSSTAVSLLESGLVTVAEQHPVTFVFNFRGLAELLKVSQGM
             QKRWKTNVHIVGPPGCKSKWAANFADPTTYWAPPRKNKWDGTHGEEVVYIDDFY
             GWLPWDDLRLCDRYLTVETKGTVPFLARSILITSNQTPLEWYSTAVPAVEALYR
             RITSLVFWKNATEQSTEGGQFVTLSPCPPEPEYINY"
BASE COUNT   452 a 361 c 492 g 463 t
ORIGIN

alignment_scores
Quality: 269.00 Length: 53
Ratio: 5.075 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.113

alignment_block
US-09-209-961-21 x PCAJ3185 ..
Align seg 1/1 to: PCAJ3185 from: 1 to: 1768
1 MetValPheIleIleHisLeuGlyPheLysTyrTpGlyValPheLysIleLy 17
:::|||||
1016 ATTGTTTATTATTATTCATTAGGTTTAGTGGGGGCTCTTAAAGATTAA 1065

17 sPheSerGluLeuTyrIleHisGlyTyrThrAspIleValValLeuValV 34
|||||
1066 ATTCTCTGAATTCATACATACATGTTACACGGATATTGTAGTCTGTGTCG 1115

34 alPheThrValPheGluArgSerAlaGluAlaTyrValValHisIleSer 50
|||||
1116 TATTACTGTTTTCGAACGACGATGCGGAGCCCTACGTGGTCCACATTCT 1165

51 ArgGlyLeu 53
|||||
1166 AGAGGTTTG 1174

seq_name: gb_vi:AF109399
seq_documentation_block: 1768 bp DNA circular VRL
LOCUS AF109399
DEFINITION Porcine circovirus type 2-E, complete genome.
ACCESSION AF109399
NID 94106896
VERSION AF109399.1 GI:4106896
KEYWORDS
```

```

/db_xref="GI:4106904"
/translation="MDIDHTVSDVDPHTAASHKSHQ"
983..988
complement(989..1033)
/codon_start=1
/product="ORF-11"
/protein_id="AAD03080.1"
/db_xref="PID:g4106906"
/db_xref="GI:4106906"
/translation="MNNKNHYEVIKKTQ"
1016..1363
/notes="similar to Porcine circovirus ORF-5 encoded by the
sequence presented in GenBank Accession Number AF027217."
/codon_start=1
/product="ORF-5"
/protein_id="AAD03075.1"
/db_xref="PID:g4106901"
/db_xref="GI:4106901"
/translation="MVFIIHLGFKWGVFKIFSELYIHGYTDIVLVVFTVFERSAEA
YVHISITGLSHPOLIFPVIIWLEIVINSIKRFGCEVPGVVGELGDCMAGGVVYVG
IGLGGYLYKVI"
complement(1022..1027)
complement(1034..1735)
/codon_start=1
/product="ORF-2"
/protein_id="AAD03072.1"
/db_xref="PID:g4106898"
/db_xref="GI:4106898"
/translation="MTIPRRRRRRRHRSHLQILRRRLVHPRHRYRWKKKNGI
FNALRSFTGYKATVSTPWSAVDMRLNLDFFPPGGTKNISIPFYIRYKVK
WFVPCSPITQGGVSSAILDDNFVIRKATQYDPYVNSRRHYPQPFYSRHY
FTPKPVLDSITDIFOPNKNRQNLWMRLQTSRNVDHVGLGTAFAENSKYDQDYNIRVTY
VQREFNLKQPLKP"
1524..1631
/codon_start=1
/product="ORF-10"
/protein_id="AAD03079.1"
/db_xref="PID:g4106905"
/db_xref="GI:4106905"
/translation="MSTAQEGVLTVALTVYKVRERALKMPPFLFOR"
complement(1528..1611)
/codon_start=1
/product="ORF-6"
/protein_id="AAD03076.1"
/db_xref="PID:g4106902"
/db_xref="GI:4106902"
/translation="MASSTPASPAPSDILSRLPQARPPGR"
1682..1741
/codon_start=1
/product="ORF-7"
/protein_id="AAD03077.1"
/db_xref="PID:g4106903"
/db_xref="GI:4106903"
/translation="MAAGAVSSSAETPPWIRHS"
join(1750..1768.1..13)
join(1762..1768.1..2)
/notes="putative; similar to the nonanucleotide motif of
Porcine circovirus presented in GenBank Accession Numbers
AF027217 and U49186"
448 a 362 c 495 g 463 t
ORIGIN

alignment_scores:
Quality: 267.00 Length: 53
Ratio: 5.135 Gaps: 0
Percent Similarity: 98.113 Percent Identity: 98.113

alignment_block:
US-09-209-961-21 x AF109399 ..
Align seg 1/1 to: AF109399 from: 1 to: 1768

1 MetValPheIleIleHisLeuGlyPheLysTrpGlyValPheLysIlely 17
|||||
1016 ATGGTTTTATTATTATAGGTTTAAAGTGGGGTCTTTAAGATTAA 1065
|||||
17 sPheSerGluLeuTyrlleHisGlyTyrrhrAspIleValValleuValv 34
|||||
1066 ATTCTCTGAATTGTACATACATGTTACACGGATATTGTAGCTCGTCG 1115
|||||
34 alpHeThrValPheGluArgSerAlaGluAlaTyrrValValHisIleSer 50
|||||
1116 TATTTACTGTTTTCGACGAGTCGCCGAGCGCTACGTGGTCCACATTCT 1165
|||||
51 ArgGlyLeu 53
|||||
1166 ACTGGTTTG 1174

seq_name: gb_v1:AF027217

seq_documentation_block:
LOCUS AF027217 1768 bp DNA circular VRL 14-MAY-1998
DEFINITION Porcine circovirus strain pmws PCV, complete genome.
ACCESSION AF027217
NID 92689645
VERSION AF027217.1 GI:2689645
KEYWORDS
SOURCE
ORGANISM
porcine circovirus.
VIRUSES; ssDNA viruses; Circoviridae; Circovirus.
REFERENCE
1 (bases 1 to 1768)
Hamel,A.L., Lin,L.L. and Nayar,G.P.
AUTHORS
Nucleotide sequence of porcine circovirus associated with
postweaning multisystemic wasting syndrome in pigs
J. Virol. 72 (6), 5262-5267 (1998)
JOURNAL
MEDLINE 98241172
REFERENCE
2 (bases 1 to 1768)
Hamel,A.L., Lin,L.L. and Nayar,G.P.S.
AUTHORS
Direct Submission
TITLE
Submitted (26-SEP-1997) Virology Laboratory, Veterinary Services
Branch, Manitoba Agriculture, 545 University Crescent, Winnipeg,
Manitoba R3T 5S6, Canada
JOURNAL
FEATURES
source
1. 1768
/organism="porcine circovirus"
/strain="pmws PCV"
/db_xref="taxon:46221"
/notes="both strands of seven overlapping PCR fragments
were sequenced; virus isolated from lung, lymph node,
spleen and tonsil tissue from pigs affected by post
weaning multisystemic wasting syndrome"
51..995
/notes="ORF1; similar to Rep protein encoded by
non-pathogenic PCV, GenBank Accession Number U49186;
predicted 35.8 kDa protein"
/codon_start=1
/product="Putative Rep protein"
/protein_id="AAC59462.1"
/db_xref="PID:g2689646"
/db_xref="GI:2689646"
/translation="MPSKNGRSGPQPHKRWFTLNPNSEDERKKIRELPISLDFYFI
VGEENEGRTPLHQGFANFVKQTFNKYKWLGCARCYIEAKGTQDNQKCYCKEGN
LLIECGAPRSGQKSDLSAVSTLLESGSLVTAEQHPVTFVRNFRGLAEILKVGKRM
QKRDWKTNVHVIYVPGCGKSKWAANFADPTTYKPNRKNWGYHGEYVVIDDFY
GWLFWDDLLRLCDRYPLTVTKGTVPFLARSILITSNQTPLEWYSSTAVPAVEALYR
RITSLVFWKNATEQSTEEGGQFVTLSPCPPEPEYNY"
117..125
/notes="glycosylation site"
complement(357..671)
/notes="ORF3; predicted 11.9 kDa protein"
327..332
/codon_start=1
/protein_id="AAC59464.1"
/db_xref="PID:g2689648"
/db_xref="GI:2689648"

misc_feature
polyA_signal
CDS

```

```
/translation="WVTIPVLSRMFPVCGFRVCKISSPFAFTTPRPWHDVYIGLPI
TLHHPAHPQKFSQPAEISDKRYRVLLCNGHOTPALQOCHSSRQVTPLSLRSSSTF
NK"
CDS
complement(386..565)
/note="ORF4; predicted 6.5 kDa protein"
/codon_start=1
/protein_id="AAC59465.1"
/db_xref="PID:g2689649"
/db_xref="GI:2689649"
/translation="WTCILVFSQRCIFPLTFKSSASPRKFLNTVNTGCCSAYVTRLPL
SKNLTAVDRSLRCP"
misc_feature
complement(470..478)
/note="glycosylation site"
CDS
complement(688..753)
/note="ORF8; predicted 2.3 kDa protein"
/codon_start=1
/protein_id="AAC59469.1"
/db_xref="PID:g2689653"
/db_xref="GI:2689653"
/translation="MDIDHTVSDVDPHTAASHKSHQ"
misc_feature
816..824
/note="glycosylation site"
misc_feature
906..914
/note="glycosylation site"
polyA_signal
983..988
complement(989..1033)
/note="ORF11; predicted 1.8 kDa protein"
/codon_start=1
/protein_id="AAC59472.1"
/db_xref="PID:g2689656"
/db_xref="GI:2689656"
/translation="MKNKNHYEVIKKTQ"
CDS
1016..1177
/note="ORF5; predicted 6.2 kDa protein"
/codon_start=1
/protein_id="AAC59466.1"
/db_xref="PID:g2689650"
/db_xref="GI:2689650"
/translation="WVFIFHLGFKWGVFKIKFSELYIHGYTDIVLVLVYTVFERSAEA
YVHLSRGL"
polyA_site
complement(1022..1027)
CDS
complement(1034..1735)
/note="ORF2; predicted 27.8 kDa protein"
/codon_start=1
/protein_id="AAC59463.1"
/db_xref="PID:g2689647"
/db_xref="GI:2689647"
/translation="WTPYRRYRRRRPRSHLQILRRRPWLVPVHPHRYRWRKNGI
FNRLSRFGYTKATVTPSWAVDMRFNIDDFVPGGNTKISIPFEYRIRKVK
VEFPKSPITQDGRGVGSFVILDDNFVKATALTYDPVNTSSRHTIQPFYSYHSRY
FTPFPVLDSTIDYFQPNKRTQLWLRLQTSRNVHDVHGLGTAFENSIIYDQDNIIRVTM
VQRFENLADPPLKP"
misc_feature
complement(1301..1309)
/note="glycosylation site"
CDS
complement(1522..1611)
/note="ORF6; predicted 3.1 kDa protein"
/codon_start=1
/protein_id="AAC59467.1"
/db_xref="PID:g2689651"
/db_xref="GI:2689651"
/translation="MASSIPASPADSLRLSPQSRPPCRWT"
CDS
1524..1631
/note="ORF10; predicted 4.1 kDa protein"
/codon_start=1
/protein_id="AAC59471.1"
/db_xref="PID:g2689655"
/db_xref="GI:2689655"
/translation="MSTAEGVLTVVALTYPKVRERRVLKMPFFLLQR"
misc_feature
1682..1741
/note="ORF7; predicted 1.9 kDa protein"
CDS
complement(1741..1741)
/protein_id="AAC59468.1"
/db_xref="PID:g2689652"
```

```
/db_xref="GI:2689652"
/translation="MAAGAVSSAVTPPWIRHS"
complement(join(1732..1768,1..92))
/note="ORF9; predicted 4.6 kDa protein"
/codon_start=1
/protein_id="AAC59470.1"
/db_xref="PID:g2689654"
/db_xref="GI:2689654"
/translation="MWLGSASSILLAGHVAEVLPRCCRCRSALVILTAHFFRFQL"
join(1746..1768,1..13)
stem_loop
/function="putative replication site"
rep_origin
join(1762..1768,1..2)
/note="AAGTATTAC" is similar to the nonnucleotide motif
in the non-pathogenic PCV, GenBank Accession Number
U49186"
BASE COUNT 452 a 360 c 495 g 461 t
ORIGIN

alignment_scores:
Quality: 266.00 Length: 53
Ratio: 5.019 Caps: 0
Percent Similarity: 100.000 Percent Identity: 96.226

alignment_block:
US-09-209-961-21 x AF027217 ..
Align seg 1/1 to: AF027217 from: 1 to: 1768

1 MetValPheIleIleHisLeuGlyPheLysTrpGlyValPheLysIleIly 17
1016 ATGGTATTTTATTTTCATTAGGTTTAAAGTGGGGGCTCTTTAAGATTAA 1065

17 spheSerGluLeuTyrlleHisGlyTyThrAspIleValValValVal 34
1066 ATTCCTGATTTGTACATACATGTTACACGGATATTGTAGTCCTGGTGG 1115

34 alpheThrValPheGluArgSerAlaGluAlaTyValValHisIleSer 50
1116 TATATACATGTTTTCAGACGACAGTCCGAGGCTACGTGTCACCATTTCT 1165

51 ArgGlyLeu 53
1166 AGAGGTTTG 1174

seq_name: gb_vi:AF147751

seq_documentation_block:
LOCUS AF147751 1768 bp DNA VRL 08-JUN-1999
DEFINITION Porcine circovirus type 2 unknown genes.
ACCESSION AF147751
NID 95007010
VERSION AF147751.1 GI:5007010
SOURCE
ORGANISM
porcine circovirus type 2.
porcine circovirus type 2
VIRUSES; ssDNA viruses; Circoviridae; Circovirus.
REFERENCE 1 (bases 1 to 1768)
AUTHORS Pogranichnyy,R., Yoon,K.-J., Harms,P., Swenson,S., Zimmerman,J. and
Sorden,S.
TITLE Characterization of clinical and immune responses in young swine to
experimental porcine circovirus type II infection
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1768)
AUTHORS Yoon,K.-J. and Pogranichnyy,R.M.
TITLE Direct Submission
JOURNAL Submitted (30-APR-1999) Veterinary Diagnostic Laboratory, Iowa
State University, College of Veterinary Medicine, Ames, Iowa 50011,
USA
FEATURES
source
Location/Qualifiers
1..1768
/organism="porcine circovirus type 2"
/isolate="ISUVDL 98-15237"
```

```

/specific_host="Sus scrofa"
/db_xref="taxon:85708"
/note="isolated from pigs with postweaning multisystemic
wasting syndrome"
51..995
CDS
/note="ORF1; 35-36kD"
/codon_start=1
/product="unknown"
/protein_id="AAD37776.1"
/db_xref="PID:g5007011"
/db_xref="GI:5007011"
/translation="MPSKNGRSGPQHRKRWFTLNNPSEDERKKIRELPISLDFYFI
VGEENEGRTPLHQFANFVKOTENKVKWILGARCHIEKAKGTQQONKEYCSKEGN
LLIECAPSQSQORSLSTAVSTLLESGSLVVAEQHPVTFVNFRLGLAELLKVSQKM
OKRDKNTVHVAVGPPCGCKSKWAANFADPETTYMKPRNKWMDGYHGEVVVDDFY
GWLPRDDLRLICDRPLTVETKGTVPFLARSILITSNQTPLEWYSSTAVPAVEALYR
RITSLVFNKATEQSTEGGQFVTLSPPCPEFPYEINY"
complement(1034..1735)
/note="ORF2; 27-28kD"
/codon_start=1
/product="unknown"
/protein_id="AAD37777.1"
/db_xref="PID:g5007012"
/db_xref="GI:5007012"
/translation="MTYPRRRYRRRRPRSHLQILRRRPWLVPVHRRYRWRKNGI
ENRLSRRTGYVTKATVTPSWAVDMRFNIDDFVPGGNTNKISIPPEYVIRKVK
VEFWPCSPITQDRGVGTAVILDDNFVKATALTVDPPYVNSRHTIPOPFSCHSRY
FTPKPVLDSITIDFQPNRNRLQWLRLQISRNVDHVLGLTAFENSIYDQINIRVTMI
VQVREFNLKDPPLKP"
BASE COUNT      450 a      362 c      494 g      462 t
ORIGIN
```

```

alignment_scores:
  Quality: 265.00      Length: 53
  Ratio: 5.096        Gaps: 0
  Percent Similarity: 98.113  Percent Identity: 96.226

alignment_block:
US-09-209-961-21 x AF147751 ..
Align seg 1/1 to: AF147751 from: 1 to: 1768

1 MetValPheIleIleHisLeuGlyPheIlyStrpGlyValPheLysIleIy 17
|||||
1016 ATGGTTTATATTATTCATTAGGTTTAAAGTGGGGGCTTTAAGATTAA 1065

17 sPheSerGluLeuTyriIleHisGlyTyThrAspIleValValLeuValv 34
|||||
1066 ATTCCTGATGTACATACATAGGTTACACGATATTGTAGTCTCGGTCG 1115

34 alPheThrValPheGluArgSerAlaGluAlaTyrValValHisIleSer 50
|||||
1116 TATATACGTGTTTCGAACGCGGTGCGGAGGCTACGTGGTCCACATTTC 1165

51 ArgGlyLeu 53
|||||
1166 AGAGGTTTG 1174
```



OM of: US-09-209-961-21 to: N\_Geneseq\_36:\* out\_format : pfs  
Date: Dec 28, 1999 1:50 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:  
-MODEL=framet\_p2n.model -DEV=xlp  
-Q/cn2\_1/USPTO\_spool/US09209961/runat\_22121999\_101043\_17942/app\_query.fasta.1  
-DB=N\_Geneseq\_36 -Qfmt=fastap -SUFFIX=ring -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-GAPOP=6.000 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=biosum62  
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR\_SCORE=escore  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09209961  
-NCPU=6 -ICPU=3 -NO\_XLPXY -WAIT -THREADS=1

Search information block:  
Query: US-09-209-961-21  
Query length: 53  
Database: N\_Geneseq\_36:\*  
Database sequences: 311585  
Database length: 125096042  
Search time (sec): 140.030000

score\_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
N_Geneseq_36:X35213	+	270.00	8.8e-31	1768	Nucleotide sequence of PCV iso
N_Geneseq_36:X35012	+	270.00	2.4e-30	1768	Genomic DNA sequence of PCV iso
N_Geneseq_36:X35212	+	270.00	2.4e-30	1768	Nucleotide sequence of PCV iso
N_Geneseq_36:X35214	+	270.00	2.4e-30	1768	Nucleotide sequence of PCV iso
N_Geneseq_36:X35210	+	270.00	524.27	1767	Nucleotide sequence of PCV iso
N_Geneseq_36:X35211	+	210.00	524.27	1767	Nucleotide sequence of PCV iso
N_Geneseq_36:X35013	-	68.00	152.71	1759	Nucleotide sequence which has
N_Geneseq_36:X35013	-	58.00	132.16	992	Babesia equi probe 10. Nucleoti
N_Geneseq_36:X35218	+	47.00	131.58	9.96	Delta thioredoxin sequence PCR
N_Geneseq_36:X35219	+	54.00	131.19	10.47	Human secreted protein 5' EST
N_Geneseq_36:X35216	+	50.00	129.78	12.54	Sugar beet acidic chitinase SE
N_Geneseq_36:X35217	+	67.00	125.45	21.85	Ocean pout DNA sequence in opa
N_Geneseq_36:X35218	+	56.50	124.63	24.29	Homo sapiens secreted protein
N_Geneseq_36:X35219	+	50.00	124.52	24.64	Human gene signature HUMGS02832
N_Geneseq_36:X35210	-	55.50	123.87	26.76	Ehrlichia sp. HGE-2 3'-end DNA
N_Geneseq_36:X35211	-	49.50	123.25	29.00	Human biallelic polymorphic DNA
N_Geneseq_36:X35212	-	53.50	122.78	30.78	Chimeric Soybean albumin 1/3 co
N_Geneseq_36:X35213	-	51.00	122.75	30.89	Staphylococcus aureus contig SE
N_Geneseq_36:X35214	-	59.00	122.35	32.54	Clone GP3 encoding major surf
N_Geneseq_36:X35215	+	50.50	121.37	36.87	Enterococcus faecalis genome co
N_Geneseq_36:X35216	+	46.00	121.08	38.27	Human biallelic polymorphic DNA
N_Geneseq_36:X35217	+	50.00	120.61	40.64	ACNPV ORF 87, residues 74356-74
N_Geneseq_36:X35218	+	52.00	120.31	42.26	Human tumour suppressor TS10q23
N_Geneseq_36:X35219	+	45.00	119.90	44.53	3' end fragment of VJcns-Bu-NP
N_Geneseq_36:X35220	+	47.00	118.66	52.24	Sequence of intermedate plasm
N_Geneseq_36:X35221	+	49.50	118.61	52.55	Staphylococcus aureus contig SE
N_Geneseq_36:X35222	+	55.50	118.12	55.93	Ehrlichia sp. extended HGE-2
N_Geneseq_36:X35223	+	52.00	117.96	57.10	Streptococcus pneumoniae genom
N_Geneseq_36:X35224	+	50.50	117.77	58.56	Hepatitis B core gene. Recombin
N_Geneseq_36:X35225	+	45.50	117.73	58.85	Interleukin 6 non-risk genotype
N_Geneseq_36:X35226	+	47.50	116.45	69.32	Human brain Expressed Sequence
N_Geneseq_36:X35227	+	47.00	116.25	71.10	EST clone EY197. New polynucle
N_Geneseq_36:X35228	+	47.00	115.93	74.13	Delta thioredoxin sequence. Pro
N_Geneseq_36:X35229	+	48.00	115.62	77.16	EST clone GP274. New polynucle
N_Geneseq_36:X35230	+	49.00	115.57	77.62	Borrelia burgdorferi polynucle
N_Geneseq_36:X35231	+	42.00	115.24	80.93	DNA polymerase ligand to a therm
N_Geneseq_36:X35232	-	49.00	115.19	81.44	H. pylori cytoplasmic protein c
N_Geneseq_36:X35233	+	45.00	115.00	83.52	Human genome fragment. New nucl
N_Geneseq_36:X35234	+	50.50	114.93	84.25	Soybean albumin 3 coding sequen
N_Geneseq_36:X35235	+	45.00	114.40	90.17	EST clone GN82. New polynucleot
N_Geneseq_36:X35236	+	47.00	114.08	93.89	Human genome fragment. (Prefer
N_Geneseq_36:X35237	-	47.50	113.97	95.29	EST clone CT207. New polynucle

N\_Geneseq\_36:V22741 + 52.00 113.90 96.10 1287 ! Babesia microti BMNI-11 ant  
N\_Geneseq\_36:T91241 - 39.00 113.70 98.60 41 ! Pseudomonas fluorescens lipas  
N\_Geneseq\_36:X33156 - 48.00 113.43 102.17 465 ! Potato isoamylase type DBE 3

seq\_name: N\_Geneseq\_36:X35213

seq\_documentation\_block:

ID X35213 standard; DNA; 1768 BP.

AC X35213, 1999 (first entry)

DE Nucleotide sequence of PCV isolate Impl010.

KW PCV isolate; type II porcine circovirus; PCV; PMWS;

OS porcine multisystemic wasting syndrome; pig; vaccine; ss.

QW Porcine circovirus.

PN FR2769322-AL.

PD 09-APR-1999.

PF 22-JAN-1998; 000873.

PR 03-OCT-1997; FR-012382.

PA (MERI-) MERIAL SAS.

PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,

PI Haines D, Harding J, Hassard L, Meehan B;

DR WPI; 99-246948/21.

PT New type II porcine circovirus, used for, e.g. passive immunization

PS of pregnant sows

PS Claim 14: Fig 4: 48pp; French.

CC The present sequence represents the nucleotide sequence of PCV isolate

CC Impl010. The specification describes a preparation of type II

CC porcine circovirus (PCV), which is particularly isolated from a lesion,

CC from a pig with symptoms of PMWS (porcine multisystemic wasting

CC syndrome). PCV (attenuated or inactivated), polypeptides derived from

CC it, and vectors that express these polypeptides are all useful in

CC vaccines, suitable for administration to adult or young pigs, or to

CC pregnant sows (for passive immunization of their offspring). DNA

CC isolated from PCV is used for in vivo or in vitro expression of viral

CC polypeptides, also as probes or primers for diagnosis in usual

CC hybridization or amplification assays. These polypeptides may also be

CC used diagnostically to detect PCV-specific antibodies, while antibodies

CC raised against the polypeptides can be used to detect antigens, in any

CC usual immunoassay format.

SQ Sequence 1768 BP; 451 A; 361 C; 495 G; 461 T;

alignment\_scores:

Quality: 273.00 Length: 53

Ratio: 5.151 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-209-961-21 x X35213 ..

Align seg 1/1 to: X35213 from: 1 to: 1768

1 MetValPheIleIleHisLeuGlyPheLysTrpGlyValPheLysIlely 17

1363 ATGGTTTATTATTATTATTAGTTTACGGGTTTACGGGTTTAAAGATTAA 1412

17 sPheSerGluLeuTyIleHisGlyTyThrAspIleValValLeuVal 34

1413 ATTCCTGAATGTACATACATAGTTTACACGGATATTAGTCTGTGTCG 1462

34 alpheThrValPheGluArgSerAlaGluAlaTyThrValValHisIleSer 50

1463 TATTACTGTTTTCGAACACGAGCGCGGAGGCTACGTGGTCCACATTTC 1512

51 ArgGlyLeu 53

1513 AGAGGTTTG 1521

seq\_name: N\_Geneseq\_36:X35012

seq\_documentation\_block:

ID X35012 standard; DNA; 1768 BP.

AC X35012;

DT 01-JUL-1999 (first entry)  
 DE Genomic DNA sequence of PCV strain 999PCV.  
 KW PCV; strain PCV999; pig; PMWS; porcine multisystemic wasting syndrome;  
 KW vaccine; ss.  
 OS Porcine circovirus  
 PN FR2769321-Al.  
 PD 09-APR-1999.  
 PF 03-OCT-1997; 012382.  
 PR 03-OCT-1997; FR-012382.  
 PA (MERI-) MERIAL SAS.  
 PA (UYBE-) UNIV QUEENS BELFAST.  
 PA (UYSA-) UNIV SASKATCHEWAN.  
 PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,  
 PI Haines D, Harding J, Hassard L, Meehan B;  
 DR WPI: 99-246948/21.  
 DR New porcine circovirus from animals with porcine systemic wasting  
 PT syndrome  
 PS Claim 13; Fig 1; 35pp; French.  
 CC The present sequence represents the genomic sequence of porcine  
 CC circovirus (PCV) strain PCV999. The sequence is obtained from PCV  
 CC isolated from a physiological or tissue sample, particularly from  
 CC a lesion, from a pig showing symptoms of PMWS (porcine multisystemic  
 CC wasting syndrome), or cultured cells, infected with PCV isolated from  
 CC such pigs. PCV (attenuated or inactivated), polypeptides derived from  
 CC it, and vectors that express these polypeptides are all useful in  
 CC vaccines, suitable for administration to adult or young pigs, or to  
 CC pregnant sows (for passive immunization of their offspring). DNA  
 CC isolated from PCV is used for in vivo or in vitro expression of viral  
 CC polypeptides, also as probes or primers for diagnosis in usual  
 CC hybridization or amplification assays. These polypeptides may also be  
 CC used diagnostically to detect PCV-specific antibodies, while antibodies  
 CC raised against the polypeptides can be used to detect antigens, in any  
 CC usual immunoassay format.  
 SQ Sequence 1768 BP; 453 A; 359 C; 486 G; 455 T;  
  
 alignment\_scores:  
 Quality: 270.00 Length: 53  
 Ratio: 5.094 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 98.113  
  
 alignment\_block:  
 US-09-209-961-21 x X35012 ..  
 Align seg 1/1 to: X35012 from: 1 to: 1768  
  
 1 MetValPheIleIleHisLeuGlyPheLysTrpGlyValPheLysIleLy 17  
 |||||  
 1364 ATGGTTTTTATTTCATTATTTAGGTTTAAAGTGGGGGCTTTAAGATTAA 1413  
  
 17 sPheSerGluLeuTyrIleHisGlyTyrThrAspIleValValLeuValV 34  
 |||||  
 1414 ATTCCTCGAATTGTACATACATGTTACACGGATATTGTAGTCTGGTCG 1463  
  
 34 alPheThrValPheGluArgSerAlaGluAlaTyrValValHisIleSer 50  
 |||||  
 1464 TATATACTGTTTTCGAACGACGAGTCCGAGGCTACGTGGTCCACATTTC 1513  
  
 51 ArgGlyLeu 53  
 |||||  
 1514 AGAGTTTG 1522  
  
 seq\_name: N\_Geneseq\_36:X35212  
  
 seq\_documentation\_block:  
 ID X35212 standard; DNA; 1768 BP.  
 AC X35212;  
 DE 01-JUL-1999 (first entry)  
 DE Nucleotide sequence of PCV isolate Imp999 (corrected version).  
 KW PCV isolate; type II porcine circovirus; PCV; PMWS;  
 KW porcine multisystemic wasting syndrome; pig; vaccine; ss.  
 OS Porcine circovirus.  
 PN FR2769322-Al.  
 PD 09-APR-1999.  
 PF 03-OCT-1997; 000873.  
 PR 03-OCT-1997; FR-012382.  
 PA (MERI-) MERIAL SAS.  
 PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,  
 PI Haines D, Harding J, Hassard L, Meehan B;  
 DR WPI: 99-246948/21.  
 DR New type II porcine circovirus, used for, e.g. passive immunization  
 PT of pregnant sows

PD 09-APR-1999.  
 PF 22-JAN-1998; 000873.  
 PR 03-OCT-1997; FR-012382.  
 PA (MERI-) MERIAL SAS.  
 PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,  
 PI Haines D, Harding J, Hassard L, Meehan B;  
 DR WPI: 99-246948/21.  
 DR New type II porcine circovirus, used for, e.g. passive immunization  
 PT of pregnant sows  
 PS Claim 14; Fig 3; 48pp; French.  
 CC The present sequence represents the nucleotide sequence of PCV isolate  
 CC Imp999. The specification describes a preparation of type II  
 CC porcine circovirus (PCV), which is particularly isolated from a lesion,  
 CC from a pig with symptoms of PMWS (porcine multisystemic wasting  
 CC syndrome). PCV (attenuated or inactivated), polypeptides derived from  
 CC it, and vectors that express these polypeptides are all useful in  
 CC vaccines, suitable for administration to adult or young pigs, or to  
 CC pregnant sows (for passive immunization of their offspring). DNA  
 CC isolated from PCV is used for in vivo or in vitro expression of viral  
 CC polypeptides, also as probes or primers for diagnosis in usual  
 CC hybridization or amplification assays. These polypeptides may also be  
 CC used diagnostically to detect PCV-specific antibodies, while antibodies  
 CC raised against the polypeptides can be used to detect antigens, in any  
 CC usual immunoassay format.  
 SQ Sequence 1768 BP; 453 A; 360 C; 494 G; 461 T;  
  
 alignment\_scores:  
 Quality: 270.00 Length: 53  
 Ratio: 5.094 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 98.113  
  
 alignment\_block:  
 US-09-209-961-21 x X35212 ..  
 Align seg 1/1 to: X35212 from: 1 to: 1768  
  
 1 MetValPheIleIleHisLeuGlyPheLysTrpGlyValPheLysIleLy 17  
 |||||  
 1363 ATGGTTTTTATTTCATTATTTAGGTTTAAAGTGGGGGCTTTAAGATTAA 1412  
  
 17 sPheSerGluLeuTyrIleHisGlyTyrThrAspIleValValLeuValV 34  
 |||||  
 1413 ATTCCTCGAATTGTACATACATGTTACACGGATATTGTAGTCTGGTCG 1462  
  
 34 alPheThrValPheGluArgSerAlaGluAlaTyrValValHisIleSer 50  
 |||||  
 1463 TATATACTGTTTTCGAACGACGAGTCCGAGGCTACGTGGTCCACATTTC 1512  
  
 51 ArgGlyLeu 53  
 |||||  
 1513 AGAGTTTG 1521  
  
 seq\_name: N\_Geneseq\_36:X35214  
  
 seq\_documentation\_block:  
 ID X35214 standard; DNA; 1768 BP.  
 AC X35214;  
 DE 01-JUL-1999 (first entry)  
 DE Nucleotide sequence of PCV isolate Imp999 (uncorrected version).  
 KW PCV isolate; type II porcine circovirus; PCV; PMWS;  
 KW porcine multisystemic wasting syndrome; pig; vaccine; ss.  
 OS Porcine circovirus.  
 PN FR2769322-Al.  
 PD 09-APR-1999.  
 PF 22-JAN-1998; 000873.  
 PR 03-OCT-1997; FR-012382.  
 PA (MERI-) MERIAL SAS.  
 PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,  
 PI Haines D, Harding J, Hassard L, Meehan B;  
 DR WPI: 99-246948/21.  
 DR New type II porcine circovirus, used for, e.g. passive immunization  
 PT of pregnant sows

PS Claim 14; Fig 6; 48pp; French.  
 CC The present sequence represents the nucleotide sequence of PCV isolate  
 CC Imp999. The specification describes a preparation of type II  
 CC porcine circovirus (PCV), which is particularly isolated from a lesion,  
 CC from a pig with symptoms of PMWS (porcine multisystemic wasting  
 CC syndrome). PCV (attenuated or inactivated), polypeptides derived from  
 CC it, and vectors that express these polypeptides are all useful in  
 CC vaccines, suitable for administration to adult or young pigs, or to  
 CC pregnant sows (for passive immunization of their offspring). DNA  
 CC isolated from PCV is used for in vivo or in vitro expression of viral  
 CC polypeptides, also as probes or primers for diagnosis in usual  
 CC hybridization or amplification assays. These polypeptides may also be  
 CC used diagnostically to detect PCV-specific antibodies, while antibodies  
 CC raised against the polypeptides can be used to detect antigens, in any  
 CC usual immunoassay format.  
 SQ Sequence 1768 BP; 453 A; 359 C; 486 G; 455 T;

alignment\_scores:  
 Quality: 270.00 Length: 53  
 Ratio: 5.094 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 98.113

alignment\_block:  
 US-09-209-961-21 x X35214 ..

Align seg 1/1 to: X35214 from: 1 to: 1768

1 MetValPheIleIleHisLeuGlyPheTyrTrpGlyValPheLysIleLy 17  
 |||||  
 1364 ATGGTTTATTATTCATTTAGTGGGGGCTTTAAGATTAA 1413  
 17 sPheSerGluLeuTyrIleHisGlyTyrThrAspIleValValLeuValV 34  
 |||||  
 1414 ATTCTCTGAATTGTACATACATGTTACACGGATATTGATCTCGTGC 1463  
 34 alPheThrValPheGluArgSerAlaGluAlaTyrValValHisIleSer 50  
 |||||  
 1464 TATATACTGTTTTCGAACGACGTCCGAGGCTACGTGTCACATTCT 1513

51 ArgGlyLeu 53  
 |||||  
 1514 AGAGGTTTG 1522

seq\_name: N\_Geneseq\_36:X35210

seq\_documentation\_block:  
 AC X35210 standard; DNA; 1767 BP.  
 AC X35210; Fig 2; 48pp; French.  
 DE Nucleotide sequence of PCV isolate Impl011-48121.  
 KW PCV isolate; type II porcine circovirus; PCV; PMWS;  
 KW porcine multisystemic wasting syndrome; pig; vaccine; ss.  
 OS Porcine circovirus.  
 PN FR2769322-Al.  
 PD 09-APR-1999.  
 PF 22-JAN-1998; 000873.  
 PR 03-OCT-1997; FR-012382.

PA (MERI-) Merial SAS.  
 PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,  
 PI Haines D, Harding J, Hassard L, Meehan B;  
 DR WPI; 99-246948/21.  
 PT New type II porcine circovirus, used for, e.g. passive immunization  
 of pregnant sows  
 PS Claim 14; Fig 1; 48pp; French.  
 CC The present sequence represents the nucleotide sequence of PCV isolate  
 CC Impl011-48121. The specification describes a preparation of type II  
 CC porcine circovirus (PCV), which is particularly isolated from a lesion,  
 CC from a pig with symptoms of PMWS (porcine multisystemic wasting  
 CC syndrome). PCV (attenuated or inactivated), polypeptides derived from  
 CC it, and vectors that express these polypeptides are all useful in  
 CC vaccines, suitable for administration to adult or young pigs, or to  
 CC pregnant sows (for passive immunization of their offspring). DNA

CC isolated from PCV is used for in vivo or in vitro expression of viral  
 CC polypeptides, also as probes or primers for diagnosis in usual  
 CC hybridization or amplification assays. These polypeptides may also be  
 CC used diagnostically to detect PCV-specific antibodies, while antibodies  
 CC raised against the polypeptides can be used to detect antigens, in any  
 CC usual immunoassay format.  
 SQ Sequence 1767 BP; 447 A; 360 C; 502 G; 458 T;

alignment\_scores:  
 Quality: 210.00 Length: 53  
 Ratio: 4.200 Gaps: 0  
 Percent Similarity: 94.340 Percent Identity: 86.792

alignment\_block:  
 US-09-209-961-21 x X35210 ..

Align seg 1/1 to: X35210 from: 1 to: 1767

1 MetValPheIleIleHisLeuGlyPheLysTrpGlyValPheLysIleLy 17  
 |||||  
 1363 ATGGTTTATTATTCATTAAGGTT.AAGTGGGGGCTTTAAGATTAA 1411  
 17 sPheSerGluLeuTyrIleHisGlyTyrThrAspIleValValLeuValV 34  
 |||||  
 1412 ATTCTCTGAATTGTACATACATGTTACACGGATATTGATCTCGTGC 1461  
 34 alPheThrValPheGluArgSerAlaGluAlaTyrValValHisIleSer 50  
 |||||  
 1462 TATATACTGTTTTCGAACGACGTCCGAGGCTACGTGTCACATTCTC 1511

51 ArgGlyLeu 53  
 |||||

1512 AGCAGTTTG 1520

seq\_name: N\_Geneseq\_36:X35211

seq\_documentation\_block:  
 ID X35211 standard; DNA; 1767 BP.  
 AC X35211;  
 DT 01-JUL-1999 (first entry)  
 DE Nucleotide sequence of PCV isolate Impl011-48285.  
 KW PCV isolate; type II porcine circovirus; PCV; PMWS;  
 KW porcine multisystemic wasting syndrome; pig; vaccine; ss.  
 OS Porcine circovirus.  
 PN FR2769322-Al.  
 PD 09-APR-1999.  
 PF 22-JAN-1998; 000873.  
 PR 03-OCT-1997; FR-012382.

PA (MERI-) Merial SAS.  
 PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,  
 PI Haines D, Harding J, Hassard L, Meehan B;  
 DR WPI; 99-246948/21.  
 PT New type II porcine circovirus, used for, e.g. passive immunization  
 of pregnant sows  
 PS Claim 14; Fig 2; 48pp; French.  
 CC The present sequence represents the nucleotide sequence of PCV isolate  
 CC Impl011-48285. The specification describes a preparation of type II  
 CC porcine circovirus (PCV), which is particularly isolated from a lesion,  
 CC from a pig with symptoms of PMWS (porcine multisystemic wasting  
 CC syndrome). PCV (attenuated or inactivated), polypeptides derived from  
 CC it, and vectors that express these polypeptides are all useful in  
 CC vaccines, suitable for administration to adult or young pigs, or to  
 CC pregnant sows (for passive immunization of their offspring). DNA  
 CC isolated from PCV is used for in vivo or in vitro expression of viral  
 CC polypeptides, also as probes or primers for diagnosis in usual  
 CC hybridization or amplification assays. These polypeptides may also be  
 CC used diagnostically to detect PCV-specific antibodies, while antibodies  
 CC raised against the polypeptides can be used to detect antigens, in any  
 CC usual immunoassay format.  
 SQ Sequence 1767 BP; 448 A; 359 C; 500 G; 460 T;

alignment\_scores:  
 Quality: 210.00 Length: 53  
 Ratio: 4.200 Gaps: 0  
 Percent Similarity: 94.340 Percent Identity: 86.792

alignment\_block:  
 US-09-209-961-21 x X35211 ..

Align seg 1/1 to: X35211 from: 1 to: 1767

```

1 MetValPheIleHisLeuGlyPheLysTrpGlyValPheLysIleLy 17
|||||
1363 ATGCTTTTATTATTCAATTAAGGTT.AAGTGGGGGCTTTAAGATTAA 1411

17 sPheSerGluLeuTyrlleHisGlyTyThrAspIleValValLeuValv 34
|||||
1412 ATTCTGAATGTACATACATAGTTACCGATATTATTCTCGGTGCG 1461

34 alpheThrValPheGluArgSerAlaGluAlaTyThrValValHisIleSer 50
|||||
1462 TATATACGTGTTTCGAAGCGAGTCGCGAGGCTAGGTGCTACATTTC 1511

51 ArgGlyLeu 53
|||||
1512 AGTAGTTTG 1520

```

seq\_name: N\_Geneseq\_36.X35013

seq\_documentation\_block:  
 ID X35013 standard; DNA; 1759 BP.  
 AC X35013;  
 DT 01-JUL-1999 (first entry)  
 DE Nucleotide sequence which has homology to PCV sequence.  
 KW PCV; strain PCV999; pig; PMWS; porcine multisystemic wasting syndrome;  
 KW vaccine; ss.  
 OS Sus sp.  
 PN FR2769321-AL.  
 PD 09-APR-1999.  
 PF 03-OCT-1997; 012382.  
 PR 03-OCT-1997; FR-012382.  
 PA (NERI-) MERIAL SAS.  
 PA (UYBE-) UNIV QUEENS BELFAST.  
 PA (UYSA-) UNIV SASKATCHEWAN.  
 PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,  
 PI Haines D, Harding J, Hassard L, Meehan B;  
 DR WPI; 99-246947/21.  
 PT New porcine circovirus from animals with porcine systemic wasting  
 PT syndrome  
 PS Disclosure; Fig 2; 35pp; French.  
 CC The specification describes a genomic sequence of porcine  
 CC circovirus (PCV) strain PCV999. The sequence is obtained from PCV  
 CC isolated from a physiological or tissue sample, particularly from  
 CC a lesion, from a pig showing symptoms of PMWS (porcine multisystemic  
 CC wasting syndrome), or cultured cells, infected with PCV isolated from  
 CC such pigs. PCV (attenuated or inactivated), polypeptides derived from  
 CC it, and vectors that express these polypeptides are all useful in  
 CC vaccines, suitable for administration to adult or young pigs, or to  
 CC pregnant sows (for passive immunization of their offspring). DNA  
 CC isolated from PCV is used for in vivo or in vitro expression of viral  
 CC polypeptides, also as probes or primers for diagnosis in usual  
 CC hybridization or amplification assays. These polypeptides may also be  
 CC used diagnostically to detect PCV-specific antibodies, while antibodies  
 CC raised against the polypeptides can be used to detect antigens, in any  
 CC usual immunoassay format. The present sequence appears in the  
 CC specification.  
 SQ Sequence 1759 BP; 475 A; 510 C; 342 G; 432 T;

alignment\_scores:  
 Quality: 68.00 Length: 41  
 Ratio: 2.519 Gaps: 0  
 Percent Similarity: 65.854 Percent Identity: 43.902

alignment\_block:  
 US-09-209-961-21 x X35013/rev ..

Align seg 1/1 to reverse of: X35013 from: 1 to: 1759

```

13 ValPheLysIleLysPheSerGluLeuTyrlleHisGlyTyThrAspI1 29
|||||
375 GTCCTTTAGGATAAATCTCTCAATTGTACATAAATAGTCAGCCTTACCA 326

29 eValValLeuValValPheThrValPheGluArgSerAlaGluAlaTyrv 46
|||||
325 CATAAATTTTGGCTGTGGCTGCATTTTGGAGCGCATAGCCGAGGCGCTGTG 276

46 alValHisIleSerArgGlyLeu 53
|||||
275 TGCTCGACATTTGGTGGGTATT 253

seq_name: N_Geneseq_36:Q34688

```

seq\_documentation\_block:  
 ID Q34688 standard; DNA; 992 BP.  
 AC Q34688; 1993 (first entry)  
 DT 10-MAY-1993 (first entry)  
 DE Babesia equi probe 10.  
 KW Probe; screen; infected animal; carrier; tick; vector; CDNA library;  
 KW plasmid; pUC13; horse; high specificity; protozoan; repetitive DNA; ss.  
 OS Babesia equi.  
 PN ZA910544S-A.  
 PD 29-APR-1992.  
 PF 12-JUL-1991; 005449.  
 PR 20-APR-1990; ZA-003002.  
 PA (TECH-) TECHNOLOGY FINANCE CORP PTY LTD.  
 PI Ambrosio RE, Posnett ES;  
 DR WPI; 93-018413/02.  
 PT Nucleotide probes for detecting Babesia equi in blood - by  
 PT hybridising with portion of genome of Babesia equi etc.  
 PS Claim 17; Fig 7; 24pp; English.  
 CC This sequence is a portion of the B. equi genome which was used to  
 CC as a probe to detect the presence of B. equi. Further probe sequences  
 CC of the invention are given in Q34678-86. These probes can be used to  
 CC screen animals for the presence of Babesia equi and to identify  
 CC infected animals and carriers of the disease. They may also be used  
 CC to detect Babesia parasites in the tick vectors. The probes were  
 CC isolated, using standard techniques from a CDNA library of B. equi  
 CC which was constructed in the plasmid vector, pUC13. The library was  
 CC screened and the clones which did not cross-hybridise to horse DNA were  
 CC investigated further. The resulting probes show high specificity for  
 CC the particular protozoan species and can be synthesised in multiple  
 CC copies. These probes are based on a repetitive DNA sequence which  
 CC occurs a number of times in the genome of B. equi.  
 SQ Sequence 992 BP; 259 A; 195 C; 197 G; 339 T;

alignment\_scores:  
 Quality: 58.00 Length: 55  
 Ratio: 1.871 Gaps: 2  
 Percent Similarity: 56.364 Percent Identity: 30.909

alignment\_block:  
 US-09-209-961-21 x Q34688 ..

Align seg 1/1 to: Q34688 from: 1 to: 992

```

3 PheIleHisLeuGlyPheLysTrpGly.....ValPh 14
|||||
485 TTGTCACTACATTAGGATTCACACTTGTGGCCTTTGCCATTATGTT 534

14 elysIleLysPheSerGluLeuTyrlleHisGly..... 25
|||||
535 TTCTGTGAGTTTTTACCAATGACCTTTTACATGCGGTAGGAAGGGGTGG 584

26 .....TyrThrAspIleValValLeuValPheThrValPheGluArg 40
|||||

```







[illegible]





```
seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-618-911-3
seq_documentation_block:
; Sequence 3, Application US/08618911
; Patent No. 5850016
; GENERAL INFORMATION:
; APPLICANT: Jung, Rudolf
; APPLICANT: Hastings, Craig
; APPLICANT: Coughlan, Sean
; APPLICANT: Hu, David
; TITLE OF INVENTION: ALTERATION OF AMINO ACID COMPOSITIONS IN
; SEEDS
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PIONEER HI-BRED INTERNATIONAL, INC.
; STREET: 700 Capital Square, 400 Locust Street
; CITY: Des Moines
; STATE: Iowa
; COUNTRY: USA
; ZIP: 50309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/618,911
; FILING DATE: Concurrently herewith
; ATTORNEY/AGENT INFORMATION:
; NAME: Simon, Soma
; REGISTRATION NUMBER: 37,444
; REFERENCE/DOCKET NUMBER: 365-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (515) 248-4800
; TELEFAX: (515) 248-4844
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 777 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 28..501
; US-08-618-911-3

alignment_scores:
Quality: 50.50 Length: 39
Ratio: 2.020 Gaps: 1
Percent Similarity: 64.103 Percent Identity: 25.641

alignment_block:
US-09-209-961-21 x US-08-618-911-3/rev ..
Align seg 1/1 to reverse of: US-08-618-911-3 from: 1 to: 777
4 IleIleHisLeuGlyPheLysTrpGlyValPheLysIleLysPheSerG1 20
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
315 CTTCTGCATGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTG 266
20 uLeuTyrlle.....HisGlyTyThrAspIleValVal 32
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
265 TCCTTATGATGTAGTCTCTCCCGCATGTCCTCGAATGATGATTGTCG 216
32 euValValPheThrVal 37
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
215 TCGTCATCATCATCATC 199

seq_name: /cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US96-09451-37
seq_documentation_block:
; Sequence 3, Application US/08464164
; Patent No. 5614195
; GENERAL INFORMATION:
; APPLICANT: Tomley, Fiona M.
; APPLICANT: Dunn, Paul P. J.
; seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-464-164-3

seq_documentation_block:
; Sequence 37, Application PC/TUS9609451
; GENERAL INFORMATION:
; APPLICANT: LARRY GOLD
; APPLICANT: SOMEDHA JAYASENA
; TITLE OF INVENTION: NUCLEIC ACID LIGAND INHIBITORS TO
; TITLE OF INVENTION: DNA POLYMERASES
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Swanson and Bratschun, L.L.C.
; STREET: 8400 East Prentice Ave., Suite 200
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/09451
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,426
; FILING DATE: 7-JUNE-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,720
; FILING DATE: 7-JUNE-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/484,557
; FILING DATE: 7-JUNE-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Barry J. Swanson
; REGISTRATION NUMBER: 33,215
; REFERENCE/DOCKET NUMBER: NEX43C/PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 793-3333
; TELEFAX: (303) 793-3433
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 78 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; PCT-US96-09451-37

alignment_scores:
Quality: 42.00 Length: 14
Ratio: 3.500 Gaps: 0
Percent Similarity: 85.714 Percent Identity: 57.143

alignment_block:
US-09-209-961-21 x PCT-US96-09451-37/rev ..
Align seg 1/1 to reverse of: PCT-US96-09451-37 from: 1 to: 78
19 SerCluLeuTyrlleHisGlyTyThrAspIleValValVal 32
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
72 AGCGAATCATACACAGAGGATACCTCTGACATCAGACTACTG 31

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-464-164-3
seq_documentation_block:
; Sequence 3, Application US/08464164
; Patent No. 5614195
; GENERAL INFORMATION:
; APPLICANT: Tomley, Fiona M.
; APPLICANT: Dunn, Paul P. J.
```

```
; APPLICANT: Bumstead, Janene M.
; APPLICANT: Vermeulen, Arno N.
; TITLE OF INVENTION: Coccidiosis poultry vaccine
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Akzo No. 5614195el Patent Department
; STREET: 1300 Piccard Drive, Suite 206
; CITY: Rockville
; STATE: Maryland
; COUNTRY: U.S.A.
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/464,164
; FILING DATE: June 2, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gormley, Mary E.
; REGISTRATION NUMBER: 34,409
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 258-5200
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 242 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: C-terminal
; ORIGINAL SOURCE:
; ORGANISM: Elmeria maxima
; STRAIN: Houghton
; DEVELOPMENTAL STAGE: Sporozoite
; IMMEDIATE SOURCE:
; LIBRARY: sporozoite cdna cloned in Lambda ZAPII
; CLONE: Em70-4, 5'end of clone
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3..242
; US-08-464-164-3

alignment_scores:
  Quality: 45.00      Length: 26
  Ratio: 2.500       Gaps: 1
  Percent Similarity: 69.231  Percent Identity: 42.308

alignment_block:
US-09-209-961-21 x US-08-464-164-3 ..
Align seg 1/1 to: US-08-464-164-3 from: 1 to: 242

25 GlyTyrThrAspIleValValLeuValVal.PheThrValPheGluArgS 41
|||
|||
117 GGTACCTTTGATGATCGCTCTCTGTTATTGACACACGGTGATTCGAAGT 166

41 erAlaGluAlaTyrValValHisIle 49
||
||
167 CCATGCAACTTCAGGTGATACATC 192

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-338-057-3

seq_documentation_block:
; Sequence 3, Application US/08338057
; Patent No. 5795741
; GENERAL INFORMATION:
; APPLICANT: Tomley, Fiona M.
```

```
; APPLICANT: Dunn, Paul P. J.
; APPLICANT: Bumstead, Janene M.
; APPLICANT: Vermeulen, Arno N.
; TITLE OF INVENTION: Coccidiosis poultry vaccine
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Organon Teknika Corporation
; STREET: 1330 Piccard Drive
; CITY: Rockville
; STATE: Maryland
; COUNTRY: U.S.A.
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/338,057
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 93.309078.9
; FILING DATE: 12-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Gormley, Mary E.
; REGISTRATION NUMBER: 34,409
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 258-5200
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 242 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: C-terminal
; ORIGINAL SOURCE:
; ORGANISM: Elmeria maxima
; STRAIN: Houghton
; DEVELOPMENTAL STAGE: Sporozoite
; IMMEDIATE SOURCE:
; LIBRARY: sporozoite cdna cloned in Lambda ZAPII
; CLONE: Em70-4, 5'end of clone
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3..242
; US-08-338-057-3

alignment_scores:
  Quality: 45.00      Length: 26
  Ratio: 2.500       Gaps: 1
  Percent Similarity: 69.231  Percent Identity: 42.308

alignment_block:
US-09-209-961-21 x US-08-338-057-3 ..
Align seg 1/1 to: US-08-338-057-3 from: 1 to: 242

25 GlyTyrThrAspIleValValLeuValVal.PheThrValPheGluArgS 41
|||
|||
117 GGTACCTTTGATGATCGCTCTCTGTTATTGACACACGGTGATTCGAAGT 166

41 erAlaGluAlaTyrValValHisIle 49
||
||
167 CCATGCAACTTCAGGTGATACATC 192

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-668-416-3

seq_documentation_block:
```

[illegible]

```

seq_name: /cgn2_6/ptodata/1/lna/5B_COMB.seq:US-08-477-559-17

seq_documentation_block:
; Sequence 17, Application US/08477559
; Patent No. 5723765
; GENERAL INFORMATION:
; APPLICANT: Oliver, Melvin J.
; APPLICANT: Quisenberry, Jerry E.
; APPLICANT: Trollinder, No. 5723765ma L. G.
; APPLICANT: Keim, Don L.
; TITLE OF INVENTION: Control of Plant Gene Expression
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rothwell, Figg, Ernst & Kurz, P.C.
; STREET: 555 Thirteenth Street, N. W.
; CITY: Washington
; STATE: D. C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,559
; FILING DATE: 01-AUG-1994
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Figg, E. Anthony
; REGISTRATION NUMBER: 27,195
; REFERENCE/DOCKET NUMBER: 1760-174A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 783-6040
; TELEFAX: (202) 783-6031
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 113 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-477-559-17

alignment_scores:
Quality: 42.00 Length: 21
Ratio: 3.000 Gaps: 0
Percent Similarity: 56.667 Percent Identity: 47.619

alignment_block:
US-09-209-961-21 x US-08-477-559-17/rev ..
Align seg 1/1 to reverse of: US-08-477-559-17 from: 1 to: 113

14 PhelysIleLysPheSerGluLeuTyrIleHisGlyTyrThrAspIleVa 30
||| : : : : : ||| ||||| : : : : : ||||| ||
84 TTCACCTCTACCTAGAGAGTCTTATACACTCTATCATCTGATAGAGT 35
30 lValLeuValVal 34
| ||| |||
34 GGGATTGTGCGTC 22

seq_name: /cgn2_6/ptodata/1/lna/5D_COMB.seq:US-08-995-161-17

seq_documentation_block:
; Sequence 17, Application US/08995161
; Patent No. 5925808
; GENERAL INFORMATION:
; APPLICANT: Oliver, Melvin J.

```

```

; APPLICANT: Quisenberry, Jerry E.
; APPLICANT: Trglinder, No. 5925808ma L. G.
; APPLICANT: Keim, Don L.
; TITLE OF INVENTION: Control of Plant Gene Expression
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rothwell, Figg, Ernst & Kurz, P.C.
; STREET: 555 Thirteenth Street, N. W.
; CITY: Washington
; STATE: D. C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/995,161
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/477,559
; FILING DATE: 01-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Figg, E. Anthony
; REGISTRATION NUMBER: 27,195
; REFERENCE/DOCKET NUMBER: 1760-174A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 783-6040
; TELEFAX: (202) 783-6031
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 113 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-995-161-17

```

```

alignment_scores:
  Quality: 42.00      Length: 21
  Ratio: 3.000       Gaps: 0
  Percent Similarity: 66.667  Percent Identity: 47.619

```

```
alignment_block:
US-09-209-961-21 x US-08-995-161-17/rev ..

```

```
Align seg 1/1 to reverse of: US-08-995-161-17 from: 1 to: 113
```

```

14 PheLysIleLysPheSerGluLeuTyrlleHisGlyTyThrAspIleVa 30
|||||..... ||| |||||..... ||||| ||
84 TTCACCTATCATCAGATAGATCTTATATACACTATCATCTAGAGAGT 35
30 lValLeuValVal 34
| ||| |||
34 GGGATTGGCGTC 22

```

```
seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-475-891A-3
```

```

seq_documentation_block:
; Sequence 3, Application US/08475891A
; Patent No. 5859339
; GENERAL INFORMATION:
; APPLICANT: Ronald, Pamela C.
; APPLICANT: Wang, Guo-Liang
; APPLICANT: Song, Wen-Yuang
; TITLE OF INVENTION: Procedures and Materials for Conferring
; TITLE OF INVENTION: Disease Resistance in Plants
; NUMBER OF SEQUENCES: 15

```

```

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,891A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/373,375
; FILING DATE: 17-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 02370-0589100S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5992 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(512..3149, 3993..4393)
; OTHER INFORMATION: /product= "RRK-B"
; OTHER INFORMATION: /note= "Xa21 Xanthomonas spp. disease
; OTHER INFORMATION: resistance gene RRR-B from rice (Oryza
; OTHER INFORMATION: sativa)"
; US-08-475-891A-3

```

```

alignment_scores:
  Quality: 56.50      Length: 55
  Ratio: 2.093       Gaps: 1
  Percent Similarity: 49.091  Percent Identity: 32.727

```

```
alignment_block:
US-09-209-961-21 x US-08-475-891A-3 ..

```

```
Align seg 1/1 to: US-08-475-891A-3 from: 1 to: 5992
```

```

4 IleIleHisLeuGlyPheLysTrpGlyValPheLysIleLysPheSerG1 20
|||||..... ||| |||||..... ||||| ||
5313 ATATGTTTCTATAGTTTAAATGTTCTGTATTTGTACCGNGGTTTCNGA 5362
20 uLeuTyrlleHis.....Glyt 26
||||| |||
5363 TCTGTACCGACATATTTCCATCAGTATTTATTCATTTCCGTTTCCGATA 5412
26 yThrAspIleValValLeuValValPheThrValPheGluArgSerAla 42
|||||..... ||| |||||..... ||||| ||
5413 TTTCGATATCGTTTTCGTTTCGACTTTACCGTTTCGATTTTCATTTC 5462
43 GluAlaTyrrValVal 47
||||| .....
5463 GAGAAAAATATGATT 5477

```

```
seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-477-559-18
```

```

seq_documentation_block:
; Sequence 18, Application US/08477559

```

```
; Patent No. 5723765
; GENERAL INFORMATION:
; APPLICANT: Oliver, Melvin J.
; APPLICANT: Quisenberry, Jerry E.
; APPLICANT: Trollander, No. 5723765ma L. G.
; APPLICANT: Keim, Don L.
; TITLE OF INVENTION: Control of Plant Gene Expression
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rothwell, Figg, Ernst & Kurz, p.c.
; STREET: 555 Thirteenth Street, N. W.
; CITY: Washington
; STATE: D. C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,559
; FILING DATE: 01-AUG-1994
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Figg, E. Anthony
; REGISTRATION NUMBER: 27,195
; REFERENCE/DOCKET NUMBER: 1760-174A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 783-6040
; TELEFAX: (202) 783-6031
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-477-559-18

alignment_scores:
  Quality: 42.00      Length: 21
  Ratio: 3.000       Gaps: 0
  Percent Similarity: 66.667   Percent Identity: 47.619

alignment_block:
US-09-209-961-21 x US-08-477-559-18 ..
Align seg 1/1 to: US-08-477-559-18 from: 1 to: 117

14 PhelysileLysPheSerGluLeuTyriIleHisGlyTyThrAspIleVa 30
||| : : : : : : : : : : : : : : : : : : : : : : : :
34 TTCACCTATCATCGATAGAGTCTTATATACACTCTATCATCTATCATGATAGT 83

30 lvalLeuValVal 34
| ||| |||
84 GGGATTGTGCGTC 96

seq_name: /cqn2_6/ptodata/1/lna/5D_COMB.seq:US-08-995-161-18

seq_documentation_block:
; Sequence 18, Application US/08995161
; Patent No. 5925808
; GENERAL INFORMATION:
; APPLICANT: Oliver, Melvin J.
; APPLICANT: Quisenberry, Jerry E.
; APPLICANT: Trollander, No. 5925808ma L. G.
; APPLICANT: Keim, Don L.
; TITLE OF INVENTION: Control of Plant Gene Expression
; NUMBER OF SEQUENCES: 26
```

```
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rothwell, Figg, Ernst & Kurz, p.c.
; STREET: 555 Thirteenth Street, N. W.
; CITY: Washington
; STATE: D. C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/995,161
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA: US 08/477,559
; APPLICATION NUMBER:
; FILING DATE: 01-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Figg, E. Anthony
; REGISTRATION NUMBER: 27,195
; REFERENCE/DOCKET NUMBER: 1760-174A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 783-6040
; TELEFAX: (202) 783-6031
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-995-161-18

alignment_scores:
  Quality: 42.00      Length: 21
  Ratio: 3.000       Gaps: 0
  Percent Similarity: 66.667   Percent Identity: 47.619

alignment_block:
US-09-209-961-21 x US-08-995-161-18 ..
Align seg 1/1 to: US-08-995-161-18 from: 1 to: 117

14 PhelysileLysPheSerGluLeuTyriIleHisGlyTyThrAspIleVa 30
||| : : : : : : : : : : : : : : : : : : : : : : : :
34 TTCACCTATCATCGATAGAGTCTTATATACACTCTATCATCTATCATGATAGT 83

30 lvalLeuValVal 34
| ||| |||
84 GGGATTGTGCGTC 96

seq_name: /cqn2_6/ptodata/1/lna/5B_COMB.seq:US-08-122-433-9

seq_documentation_block:
; Sequence 9, Application US/08122433
; Patent No. 5683985
; GENERAL INFORMATION:
; APPLICANT: Chu, Barbara C.F.
; APPLICANT: Orgel, Leslie
; TITLE OF INVENTION: OLIGODEOXYNUCLEOTIDES AND
; TITLE OF INVENTION: OLIGONUCLEOTIDES USEFUL AS DECOYS FOR PROTEINS WHICH
; TITLE OF INVENTION: SELECTIVELY BIND TO DEFINED DNA SEQUENCES
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PRETTY, SCHROEDER, BRUEGGEMANN & CLARK
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: California
```

```
;
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/122,433
; FILING DATE: 22-SEP-1993
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/687,337
; FILING DATE: 18-APR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P31 9308
; TELEPHONE: 619-546-1995
; TELEFAX: 619-546-9392
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 55 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; US-08-122-433-9

alignment_scores:
  Quality: 39.00      Length: 9
  Ratio: 4.875       Gaps: 0
  Percent Similarity: 88.889   Percent Identity: 66.667

alignment_block:
  US-09-209-961-21 x US-08-122-433-9/rev ..
  Align seg 1/1 to reverse of: US-08-122-433-9 from: 1 to: 55

  9 PhelystrpGlyValPheLysIleLys 17
  |||:|||||:||||| |||
  48 TTCGCTGGGACTTTTCARAAAGAA 22

seq_name: /cgn2.6/ptodata/1/ina/5A_COMB.seq:US-08-308-892A-16

seq_documentation_block:
; Sequence 16, Application US/08308892A
; Patent No. 5500341
; GENERAL INFORMATION:
; APPLICANT: Spears, Patricia A.
; TITLE OF INVENTION: SPECIES-SPECIFIC DETECTION OF
; TITLE OF INVENTION: MYCOBACTERIUM KANSASII
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Richard J. Rodrick, Becton Dickinson and
; COMPANY: Company
; STREET: 1 Becton Drive
; CITY: Franklin Lakes
; STATE: NJ
; COUNTRY: US
; ZIP: 07417
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/308,892A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
```

```
;
; NAME: Fugit, Donna R.
; REGISTRATION NUMBER: 32,135
; REFERENCE/DOCKET NUMBER: P-3128
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 171 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-308-892A-16

alignment_scores:
  Quality: 43.00      Length: 27
  Ratio: 2.263       Gaps: 0
  Percent Similarity: 70.370   Percent Identity: 29.630

alignment_block:
  US-09-209-961-21 x US-08-308-892A-16/rev ..
  Align seg 1/1 to reverse of: US-08-308-892A-16 from: 1 to: 171

  11 TrpGlyValPheLysIleLysPheSerGluLeuTyrlleHisGlyTyrrh 27
  |||:|||||:||||| |||
  113 TGGGCGCCGAGGCGTCGAGTTCGCCGAGGTGCTGTCGACGATATCC 64

  27 RasPleValValLeuValPheThrVal 37
  |||:|||||:||||| |||
  63 CAGGTCGCGCTGTGCGCCGCGTACTGTC 33

seq_name: /cgn2.6/ptodata/1/ina/5C_COMB.seq:US-08-481-337A-3

seq_documentation_block:
; Sequence 3, Application US/08481337A
; Patent No. 5863738
; GENERAL INFORMATION:
; APPLICANT: TEN DUKE, Peter
; APPLICANT: HELDIN, Carl-Henrik
; APPLICANT: MIYAZONO, Kohei
; APPLICANT: SAMPATH, Kuber T.
; TITLE OF INVENTION: Morphogenic Protein-Specific Cell
; TITLE OF INVENTION: Surface Receptors and Uses Therefor
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Testa, Hurwitz & Thibault
; STREET: 125 High St.
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/481,337A
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MEYERS, Thomas C.
; REGISTRATION NUMBER: 36,989
; REFERENCE/DOCKET NUMBER: CRP-097CP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2724 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
```

```
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 104..1630
; OTHER INFORMATION: /product= "Human ALK2"
US-08-481-337A-3

alignment_scores:
  Quality: 53.00      Length: 27
  Ratio: 2.524       Gaps: 0
  Percent Similarity: 77.778      Percent Identity: 40.741

alignment_block:
  US-09-209-961-21 x US-08-481-337A-3 ..
  Align seg 1/1 to: US-08-481-337A-3 from: 1 to: 2724
      20 GluLeuTyrIleHisGlyTyrThrAspIleValValLeuValValPheTh 36
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      2250 GAATTGTTTATACACAACTTTGCAAAATTATTATTACTTGTGCACCTAGT 2299

      36 rValPheGluArgSerAlaGluAlaTyrVal 46
      :|||||: :|||:|||||:
      2300 AGTTTTCACAAACTGCTTTGTGCATATGTT 2330

seq_name: /cgn2_6/ptodata/1/ina/PTUS9_COMB.seq:PCT-US95-05467-3

seq_documentation_block:
; Sequence 3, Application PC/TUS9505467
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: MORPHOGENIC PROTEIN-SPECIFIC CELL
; TITLE OF INVENTION: SURFACE RECEPTORS AND USES THEREFOR
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PATENT ADMINISTRATOR, TESTA, HURWITZ &
; ADDRESSEE: THIBEAULT
; STREET: 53 STATE STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/05467
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER, EDMUND R.
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: CRP-097PC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2724 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 104..1633
; OTHER INFORMATION: /product= "Human ALK2"
PCT-US95-05467-3

alignment_scores:
  Quality: 53.00      Length: 27
  Ratio: 2.524       Gaps: 0
  Percent Similarity: 77.778      Percent Identity: 40.741

alignment_block:
  US-09-209-961-21 x PCT-US95-05467-3 ..
  Align seg 1/1 to: PCT-US95-05467-3 from: 1 to: 2724
      20 GluLeuTyrIleHisGlyTyrThrAspIleValValLeuValValPheTh 36
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      2250 GAATTGTTTATACACAACTTTGCAAAATTATTATTACTTGTGCACCTAGT 2299

      36 rValPheGluArgSerAlaGluAlaTyrVal 46
      :|||||: :|||:|||||:
      2300 AGTTTTCACAAACTGCTTTGTGCATATGTT 2330
```







```

: ||| ... :|||:|||||:|||||:|||||:
122 GAAGGGACTAGT...GTGGCGGTAGTGTCTCAGGGAAGTGTTCAACT 76
40 rGserAlaGluATyrValValHisIleSerArg 51
|||||:|||||: ... :|||:|||||:
75 CATCCGAGAGAATGCTGTCAGCATCTCCAGAAA 41

```

seq\_name: gb\_gss9:A0620563

```

seq_documentation_block:
LOCUS A0620563 575 bp DNA GSS 16-JUN-1999
DEFINITION HS_5190_B1_C02_T7A RPCI-11 Human Male BAC Library Homo sapiens
genomic clone Plate=766 Col=3 Row=F, genomic survey sequence.
ACCESSION A0620563
NID 95082955
VERSION A0620563.1 GI:5082955
KEYWORDS GSS.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 575)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Furlong,J.,
Shaker,R., Schmidt,S., Frai-coff,R. and Hood,L.E.
TITLE Construction of a Characterized Clone Resource for Genomic
Sequencing
JOURNAL Unpublished (1998)
COMMENT
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Research Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 766 row: F column: 3
Seq primer: T7
Class: BAC ends
High quality sequence stop: 575.

```

```

FEATURES
source
1..575
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="plate=766 Col=3 Row=F"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Genomic sequence of BAC ends"
BASE COUNT 240 a 78 c 81 g 163 t 13 others
ORIGIN

```

```

alignment_scores:
Quality: 66.50 Length: 53
Ratio: 1.956 Gaps: 2
Percent Similarity: 64.151 Percent Identity: 28.302
alignment_block:
US-09-209-961-21 x A0620563/rev ..
Align seg 1/1 to reverse of: A0620563 from: 1 to: 575
4 lIelleHisLeuGlyPheLysTrpGlyValPheLysIleLysPheSerG1 20
:|||||:|||||: ||| ||| ||| ||| ||| |||
247 ATGTGTCATAGACATTTTGTGTGGTATTTTATATATGTTCTATCTCC 198
20 uLeuTyrlleHisGlyTyrrThrAspIleValLeuVal..... 33
:|||||: ||| :|||||:|||||:|||||:

```

```

197 TTTGTGG...CACATAAGTCTGACTTTGCTCTAATCATCAATCAAT 151
34 .....ValPheThrValPheGluArgSerAlaGluAlaValValVal 47
:|||||:|||||:|||||: ||| :|||||:|||||:
150 TTTGCCATATATTGTGATTTTAAATTTCTCTTAAAGTCTATTTATCC 101
48 HisIleSer 50
|||||:|||||:
100 CATGTAGCT 92
seq_name: gb_est24:A1226365
seq_documentation_block:
LOCUS A1226365 410 bp mRNA EST 29-OCT-1998
DEFINITION uf07b09.y1 Soares mouse mammary gland NMLMG Mus musculus cDNA clone
IMAGE:1510649 5', mRNA sequence.
ACCESSION A1226365
NID 93809418
VERSION A1226365.1 GI:3809418
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 410)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wyllie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Jan 17, 1998 this sequence version replaced gi:2045169.

```

```

Contact: Marra M/Mouse EST Project
WashU-HMMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:937501
Seq primer: -40RP from Gibco
High quality sequence stop: 346.

```

```

FEATURES
source
1..410
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:1510649"
/clone_lib="Soares mouse mammary gland NMLMG"
/sex="female (lactating)"
/tissue_type="mammary gland"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from mammary
gland tissue from a lactating female, and was then primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT73 vector. Library is normalized. Library
was constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 77 a 99 c 116 g 118 t
ORIGIN

```

```

alignment_scores:
Quality: 63.50 Length: 51
Ratio: 2.048 Gaps: 1
Percent Similarity: 60.784 Percent Identity: 27.451
alignment_block:

```

```

US-09-209-961-21 x A1226365
Align seg 1/1 to: A1226365 from: 1 to: 410

3 PheilleHisLeuGlyPheLysTrpGlyValPheLysIleLysPheSe 19
|||||::||| :: |||::||| ::|||::||| ::|||::|||
167 TTTCATTATCATCGGGCAGACCTTGGGTGGTGCTGCCGTCCGCAG 216

19 rGlueTyrrileHisGlyTyrr.....ThraspilevalV 31
|||||::|||::|||::|||::|||::|||::|||::|||
217 TGAGATATCTACAGGCATCTTGGCTCCGCACAAGGAAGTCCCTG 266

31 alLeuValValPheThrValPheGluAUGSerAlaGluAlaTyrrVal 47
|||||::|||::|||::|||::|||::|||::|||::|||
267 TGTACTAGTCAGTCTTTAGTTTTTCTATTGAAGACCAGTTCGCTGCTTG 316

48 His 48
|||
317 CAC 319

seq_name: gb_est31:AU060237

seq_documentation_block:
LOCUS AU060237 697 bp mRNA EST 20-MAY-1999
DEFINITION AU060237 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
discoideum cDNA clone SLA658, mRNA sequence.
ACCESSION AU060237
NID 94881341
VERSION AU060237.1 GI:4881341
KEYWORDS EST.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum.
REFERENCE 1 (bases 1 to 697)
AUTHORS Eukaryota; Dictyosteliida; Dictyostelium.
Yoshino,R., Morio,T. and Tanaka,Y.
TITLE Developmental cDNA in Dictyostelium discoideum
JOURNAL Unpublished (1997)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3187168.

Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
Email: d402hu@tsukuba.ac.jp
PROJECT = Dictyostelium discoideum cDNA project in Japan.

FEATURES
Source
1..697
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/map="15 q15(21.1)-q22; 1: 21q"
/clone="SLA658"
/clone_lib="dictyostelium discoideum SL (H.Urushihara)"
/dev_stage="slug"
BASE COUNT 262 a 107 c 98 g 230 t
ORIGIN

alignment_scores:
Quality: 65.00 Length: 43
Ratio: 2.097 Gaps: 0
Percent Similarity: 72.093 Percent Identity: 30.233

alignment_block:
US-09-209-961-21 x AU060237/rev ..
Align seg 1/1 to reverse of: AU060237 from: 1 to: 697

7 LeuGlyPheLysTrpGlyValPheLysIleLysPheSerGluLeuTyrrIL 23
|||||::|||::|||::|||::|||::|||::|||::|||
185 ATTGATTCTCTGTTCTTCTATTAAATTAATTTGCTTCAATATATT 136

23 ehisGlyTyrrThraspilevalValLeuValValPheThrValPheGluA 40
|||||::|||::|||::|||::|||::|||::|||::|||

```

189 GTGTCCTCCCAATGACTGTGCAAAATTTTGCTCTTATCTACTATTA 238

38 he.....GluArgSerAlaGluAlaTyrValVal 47

239 TTTAATATGAAGAAGGAAGTGAATCCTATATGTT 273

seq\_name: gb\_gss10:AQ736850

seq\_documentation\_block: 438 bp DNA GSS 15-JUL-1999  
 LOCUS AQ736850  
 DEFINITION HS\_2244\_B2\_H04\_T7C CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2244 Col=8 Row=P, genomic survey sequence.  
 ACCESSION AQ736850  
 NID Q5508402  
 VERSION AQ736850.1 GI:5508402  
 KEYWORDS GSS.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 438)  
 AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Furlong,J., Shaker,R., Schmidt,S., Traicoff,R. and Hood,L.E.  
 TITLE Construction of a Characterized Clone Resource for Genomic Sequencing  
 JOURNAL Unpublished (1998)  
 COMMENT

Contact: Mahairas GG, Wallace JC, Hood L  
 High Throughput Sequencing Center  
 University of Washington  
 401 Queen Anne Avenue North, Seattle, WA 98109, USA  
 Tel: (206) 616-3618  
 Fax: (206) 616-3887  
 Email: jwallace@u.washington.edu  
 Clones may be purchased from Research Genetics (info@resgen.com).  
 BAC end Web Server: http://www.htsc.washington.edu  
 Plate: 2244 row: P column: 8  
 Seq primer: T7  
 Class: BAC ends

High quality sequence stop: 438.

FEATURES  
 source

Location/Qualifiers  
 1..438  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="Plate=2244 Col=8 Row=P"  
 /clone\_lib="CIT Approved Human Genomic Sperm Library D"  
 /sex="male"  
 /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 102 a 100 c 88 g 146 t 2 others  
 ORIGIN

alignment\_scores:  
 Quality: 63.00 Length: 53  
 Ratio: 1.853 Gaps: 1  
 Percent Similarity: 64.151 Percent Identity: 24.528

alignment\_block:  
 US-09-209-961-21 x AQ736850 ..

Align seg 1/1 to: AQ736850 from: 1 to: 438

1 MetValPheIleHisLeuGlyPheLysTyrGlyValPheLysIleLeu 17

265 CTTATTTTTCATCATTTGGTCATCTNITGGGGAAGACAGATATA 314

17 sPheSerGluLeuTyrIleHisGlyTyrThrAspIleValValLeuVal 34

315 TTGTCCTCCAGCTCTCTCAGTAACCTC.....ATATTCATATCCGCA 358

34 aPheThrValPheGluArgSerAlaGluAlaTyrValValHisIleSer 50

359 TTCCTTATATTCAGGCACCTCCAAGGTGGATGTAGTCTCCCTGATG 408

51 ArgGlyLeu 53

409 AATGGAATA 417

seq\_name: gb\_gss7:AQ501328

seq\_documentation\_block: 639 bp DNA GSS 28-APR-1999  
 LOCUS AQ501328  
 DEFINITION V24C4 mTn-3xHA/lacZ Insertion Library Saccharomyces cerevisiae genomic 5', genomic survey sequence.  
 ACCESSION AQ501328  
 NID Q4706978  
 VERSION AQ501328.1 GI:4706978  
 KEYWORDS GSS.  
 SOURCE baker's yeast.  
 ORGANISM Saccharomyces cerevisiae  
 Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 REFERENCE 1 (bases 1 to 639)  
 AUTHORS Ross-Macdonald,P., Roemer,T., Coelho,P.S.R., Agarwal,S., Kumar,A., desEtages,S.A., Cheung,K.-H., Sheehan,A., Symoniatis,D., Jansen,R., Umansky,L., Heitman,M., Nelson,K., Iwasaki,H., Kanada,D., Lugo,R., Hager,K., Miller,P., Roeder,G.S. and Snyder,M.  
 TITLE Large-Scale Analysis of the Yeast Genome by Transposon Tagging and Gene Disruption  
 JOURNAL Unpublished (1999)  
 COMMENT

Contact: Kumar A  
 Michael Snyder, Dept. of Mol. Cell. and Dev. Biology  
 Yale University  
 P.O. Box 208103, New Haven, CT 06520-8103, USA  
 Tel: 203 432 9949  
 Fax: 203 432 6161  
 Email: anuj.kumar@yale.edu  
 te of mTn-3xHA/lacZ insertion.  
 Seq primer: GGCCTCTTCTTTGGAGTAC  
 Class: transposon-tagged.  
 Location/Qualifiers  
 1..639  
 /organism="Saccharomyces cerevisiae"  
 /db\_xref="taxon:4932"  
 /clone\_lib="mTn-3xHA/lacZ Insertion Library"  
 /note="Vector: PHSS6-Sal; A yeast genomic DNA library (lacking mitochondrial DNA) was prepared in PHSS6-Sal; genomic DNA was size-fractionated (DNA of roughly 2-3 kb in length) prior to cloning. This library was subsequently mutagenized with a mTn-3xHA/lacZ minitransposon containing lacZ, URA3, and tet resistance."

FEATURES  
 source

1..639  
 /organism="Saccharomyces cerevisiae"  
 /db\_xref="taxon:4932"  
 /clone\_lib="mTn-3xHA/lacZ Insertion Library"  
 /note="Vector: PHSS6-Sal; A yeast genomic DNA library (lacking mitochondrial DNA) was prepared in PHSS6-Sal; genomic DNA was size-fractionated (DNA of roughly 2-3 kb in length) prior to cloning. This library was subsequently mutagenized with a mTn-3xHA/lacZ minitransposon containing lacZ, URA3, and tet resistance."

BASE COUNT 181 a 137 c 141 g 179 t 1 others  
 ORIGIN

alignment\_scores:  
 Quality: 64.00 Length: 54  
 Ratio: 1.939 Gaps: 3  
 Percent Similarity: 61.111 Percent Identity: 35.185

alignment\_block:  
 US-09-209-961-21 x AQ501328/rev ..

Align seg 1/1 to reverse of: AQ501328 from: 1 to: 639

6 HisLeuGlyPheLys.TyrGlyValPheLysIleLysPheSerGluLeu 22

493 CACTTGGGCTTTCTTGGCAGGTGTTGCTCTATTCCTTCTTCAGTT 444

22 yrIleHisGlyTyrThrAspIleValValLeuValPheThrValPhe 38

443 TGGTCCGC...TATTCAGATTATTGGTGGTCTCTACTGTATACATCTCT 397





224 CTATTC 229

seq\_name: gb\_est25:AI320780

seq\_documentation\_block: 370 bp mRNA EST 18-DEC-1998  
LOCUS AI320780  
DEFINITION d1a12nm.r1 Neurospora crassa morning cDNA library Neurospora crassa  
CDNA clone d1a12nm 3', mRNA sequence.

ACCESSION AI320780

NID 94036762

VERSION AI320780.1 GI:4036762

KEYWORDS EST.

SOURCE Neurospora crassa.

ORGANISM

Eukaryota; Fungi; Ascomycota; Euascomycetes; Pyrenomycetes;

Sordariales; Sordariaceae; Neurospora.

1 (bases 1 to 370)

Zhu, H., Lai, H., Kupfer, D., Dunlap, J. C. and Roe, B. A.

Two Neurospora crassa EST Databases

Unpublished (1998)

On Oct 19, 1998 this sequence version replaced gi:454479.

Other ESTs: d1a12nm.fl

Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu

Department of Chemistry and Biochemistry

Advanced Center for Genome Technology, University of Oklahoma

620 Parrington Oval, Norman, OK 73019, USA

Tel: 405 325 4912

Fax: 405 325 7762

Email: broe@ou.edu

We anticipate the future release of the cDNA clones to the Fungal

Genetics Stock Center

Possible reversed clone: polyT not found

Seq primer: Universal Reverse Primer

High quality sequence stop: 323.

FEATURES

source

1..370

/organism="Neurospora crassa"

/strain="bd, frq7 A"

/db\_xref="taxon:5141"

/clone="d1a12nm"

/tissue\_type="tissue harvested following 22hr growth in

dark"

/note="Vector: pBluescript SK-; Site\_1: XbaI; Site\_2:

EcoRI; See: Bell-Pedersen, D., et al. PNAS 93:13096,1996.

5' end of cDNA cloned into XbaI site of pBluescript; 3'

end of cDNA cloned into EcoRI site of pBluescript"

113 a 104 c 91 g 62 t

BASE COUNT

ORIGIN

alignment\_scores:

Quality: 60.50

Ratio: 1.729

Percent Similarity: 50.725

Length: 69

Gaps: 3

Percent Identity: 26.087

alignment\_block:

US-09-209-961-21 x AI320780/rev ..

Align seg 1/1 to reverse of: AI320780 from: 1 to: 370

3 PhellelleHisLeuGlyPheLysTrp..... 11

|||||..... 11

254 TTCATCTCCGGTTGCTTCAATGGGTGGGATTTTGCTCTTAT 205

12 ..... GlyValPheLysIleLysPheSerCluLeuTyrIleHisG 25

|||||..... 11

204 TCTTTGGTCCCGCTTCGGCGGIC...TTCCGGACATGATACATCCATG 158

25 ly..... 26

|||||..... 11

157 AGAGCTTGAGGGCAACACGATATTCGCGATCAAGCATGCCCTCGG 108

|||||..... 11

27 ThrAspIleValLeuValPheThrValPheGluArgSerAlaGl 43

|||||..... 11

107 GTGGACCTCGTCAACATGCTTCTGCTCGTCATCGCGCTTCTGCATT 58

43 uLaTyr 45

|||||..... 11

57 GCGTTAC 51

seq\_name: gb\_gss5:AQ331923

seq\_documentation\_block:

LOCUS AQ331923

DEFINITION HS\_5002\_A2\_E08\_SP6E RPC111 Human Male BAC Library Homo sapiens

genomic clone Plate=578 Col=16 Row=I, genomic survey sequence.

ACCESSION AQ331923

NID 94128510

VERSION AQ331923.1 GI:4128510

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 378)

Mahairas, G. G., Wallace, J. C., Smith, K., Swartzell, S., Furlong, J.,

Shaker, R., Schmidt, S., Traicoff, R. and Hood, L. E.

Construction of a Characterized Clone Resource for Genomic

Sequencing

Unpublished (1998)

JOURNAL

COMMENT

Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Clones may be purchased from Research Genetics (info@resgen.com).

BAC end Web Server: http://www.htsc.washington.edu

Plate: 578 row: I column: 16

Seq primer: SP6

Class: BAC ends

High quality sequence stop: 378.

FEATURES

Location/Qualifiers

1..378

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clones="Plate=578 Col=16 Row=I"

/clone\_lib="RPC111 Human Male BAC Library"

/sex="Male"

/cell\_type="Lymphocytes"

/note="vector: pBAC3.6; RPC111 Human Male BAC Library"

144 a 50 c 47 g 137 t

BASE COUNT

ORIGIN

alignment\_scores:

Quality: 60.50

Ratio: 2.327

Percent Similarity: 57.778

Length: 45

Gaps: 1

Percent Identity: 26.667

alignment\_block:

US-09-209-961-21 x AQ331923 ..

Align seg 1/1 to: AQ331923 from: 1 to: 378

5 lleHisLeuGlyPheLysTrpGlyValPheLysIleLysPheSerCluLe 21

|||||..... 11

242 GTGCACTAACCTTCTCTCTATTTATCTCAATTAATGATGTCAGAT 291

|||||..... 11

21 utyIleHisGlyTyrThrAspIleValLeuValPheThrValP 38

|||||..... 11

292 ACAGACACACACACACGGATATATATATAGTA..... 328



5 IleHisLeuGlyPheLysTrpGlyValPheLysIleLysPheSerGluLeu 21  
 |||:::||||| ||| ||| ||| :::  
 30 ATTAACTTTGGCGATAAATGGGGTGTAGGGGATATATTCACACTCTGGG 79

```
204 TCFTTGGGTGCGCGTCTTCGGCGTC...TTCGGGACATGTACATCCATG 158
25 ly.....Tyr 26
157 AGAGCCTGAGGGCAACAACGATATTCGTCGGATGAGCATGCCGCTCTGG 108
27 ThrAspIleValLeuValPheThrValPheGluArgSerAlaG1 43
107 GTGGACCTCGTCAACATGCTTCTCTGGCTCGTCATCGCGGTTTCTGCATT 58
43 ualATyr 45
57 GGCCTAC 51
```

Sequence	Strd Orig	zScore	EScore	Len	Documentation
9b_v1:AF085695	+ 1714.00	2896.36	2.9e-153	1768	! AF085695 Porcine circovirus st
9b_v1:AF086836	+ 1710.00	2889.55	6.0e-153	1768	! AF086836 Porcine circovirus st
9b_v1:AF086834	+ 1706.00	2882.74	1.4e-152	1768	! AF086834 Porcine circovirus st
9b_v1:AF086835	+ 1701.00	2874.24	4.3e-152	1768	! AF086835 Porcine circovirus st
9b_v1:AF055391	+ 1682.00	2841.90	2.7e-150	1768	! AF055391 Porcine circovirus Ty
9b_v1:AF055392	+ 1682.00	2841.90	2.7e-150	1768	! AF055392 Porcine circovirus Ty
9b_v1:AF109397	+ 1682.00	2841.90	2.7e-150	1768	! AF109397 Bovine circovirus, co
9b_v1:AF147751	+ 1682.00	2841.90	2.7e-150	1768	! AF147751 Porcine circovirus Ty
9b_v1:PC0A3185	+ 1682.00	2841.90	2.7e-150	1768	! AJ231185 Porcine circovirus DN
9b_v1:AF166528	+ 1678.00	2835.09	6.1e-150	1768	! AF166528 Porcine circovirus ty
9b_v1:AF109399	+ 1677.00	2833.39	8.5e-150	1768	! AF109399 Porcine circovirus ty
9b_v1:AF055393	+ 1676.00	2831.70	1.0e-149	1767	! AF055393 Porcine circovirus Ty
9b_v1:AF055394	+ 1676.00	2831.70	1.0e-149	1767	! AF055394 Porcine circovirus Ty
9b_v1:AF027217	+ 1676.00	2831.69	1.0e-149	1768	! AF027217 Porcine circovirus Ty
9b_v1:AF117753	+ 1674.00	2828.29	1.6e-149	1768	! AF117753 Porcine circovirus ty
9b_v1:AF112862	+ 1671.00	2823.18	3.0e-149	1768	! AF112862 Porcine circovirus ty
9b_v1:AF154679	+ 1669.00	2819.78	4.6e-149	1768	! AF154679 Porcine circovirus, c
9b_v1:AF109398	+ 1665.00	2812.97	1.1e-148	1768	! AF109398 Porcine circovirus, c
9b_v1:AF012107	+ 1447.00	2442.03	5.1e-128	1759	! AF012107 Porcine circovirus, c
9b_v1:PC049186	+ 1447.00	2442.03	5.1e-128	1759	! U49186 Porcine circovirus, com
9b_v1:AF071879	+ 1443.00	2435.23	1.2e-127	1758	! AF071879 Porcine circovirus c
9b_v1:PC00G0EN	+ 1439.00	2428.41	2.9e-127	1759	! Y09921 Porcine circovirus comp
9b_v1:AF118095	+ 1063.00	1799.10	3.3e-92	597	! AF118095 Porcine circovirus ty
9b_v1:AF080560	+ 610.50	1017.25	1.2e-48	1993	! AF080560 Beak and feather dise
9b_v1:AF071878	+ 597.50	995.13	2.0e-47	1993	! AF071878 Beak and feather dise
9b_v1:AF135393	+ 289.00	484.94	5.2e-19	438	! AF135393 Porcine circovirus PCV
9b_v1:AF134203	+ 261.00	437.33	2.3e-16	436	! AF134203 Porcine circovirus PCV
9b_v1:BBU12586	+ 197.00	319.27	8.8e-10	1111	! U12586 Banana bunchy top virus
9b_v1:BBU12587	+ 197.00	319.13	8.9e-10	1127	! U12587 Banana bunchy top virus
9b_v1:NVX5968	+ 185.50	300.69	9.5e-09	1004	! AJ005968 faba bean necrotic ve
9b_v1:SCU16731	+ 182.00	294.56	2.1e-08	1022	! U16731 Subterranean clover stu
9b_v1:AB000921	+ 168.50	271.71	3.9e-07	1009	! AB000921 Milk vetch dwarf viru
9b_Pat:AR063451	+ 168.50	270.82	4.4e-07	1106	! AB063451 Sequence 3 from paten
9b_v1:BYTV1	+ 168.50	270.82	4.4e-07	1106	! L32166 Banana bunchy top virus
9b_v1:CFDQ	+ 167.00	266.75	7.4e-07	1291	! M29663 Coconut foliar decay vi
9b_Pat:AR063452	+ 163.00	261.55	1.4e-06	1096	! AR063452 Sequence 4 from paten
9b_v1:BYTV2	+ 163.00	261.55	1.4e-06	1095	! L32167 Banana bunchy top virus
9b_v1:U97525	+ 157.00	251.27	5.4e-05	1103	! U97525 Banana bunchy top virus
9b_Pat:AR063453	+ 156.00	249.58	6.6e-06	1091	! AR063453 Sequence 5 from paten
9b_Pat:AR010234	+ 156.00	249.55	6.7e-06	1105	! AR010234 Sequence 18 from paten
9b_v1:U02312	+ 152.50	243.69	1.4e-05	1093	! U02312 Banana bunchy top virus
9b_Pat:AR010232	+ 152.00	242.76	1.6e-05	1103	! AR010232 Sequence 16 from paten

CDS  
 1016 ..1177  
 /codon\_start=1  
 /product="P6.2"  
 /protein\_id="AAC61863.1"  
 /db\_xref="PID:g3668366"  
 /db\_xref="GI:3668366"  
 /translation="MVFIHLGFKWGVKIKFSELYIHGYTDIVLVVFTVFSRAEA  
 YVHISRGL"

CDS  
 complement(1034..1735)  
 /codon\_start=1  
 /product="P27.9"  
 /protein\_id="AAC61864.1"  
 /db\_xref="PID:g3668367"  
 /db\_xref="GI:3668367"  
 /translation="MTYPRRYRRRRHRRHRLGOLRRRRLVHPHRRYRWRKNGI  
 FNTLSRTFGYTVKRTTTPSWAIDMMRFKIDDFVPPGGTNNKISIFEYIRKKVK  
 VEFWPCSPITQCDRGVSGTAVLDDNFYTKATALTIDPYVNYSSRHTIPOPFYSHSRY  
 FTPKPVLDSTIDYQPNKRNQLRLQTSNVHDVGLGTAFENSKYDQDYNIRVTMY  
 VQFRENLKDPLEP"

BASE COUNT 463 a 362 c 481 g 462 t  
 ORIGIN

alignment\_scores:  
 Quality: 1714.00 Length: 314  
 Ratio: 5.459 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 99.682

alignment\_block:

US-09-209-961-3 x AF085695

Align seq 1/1 to: AF085695 from: 1 to: 1768

1 MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTr 17  
 51 ATGCCCGAGCAAGAAGATGGAAGAGCGGAGCCCAACACCATAAAGGTG 100  
 17 pValPheThrLeuAsnAsnProSerGluAspGluArgLysLysIleArg 34  
 101 GGTGTTCCAGCTCAATAATCTTCCGAAGACGAGCGCAAGAAATACGGG 150  
 34 LuLeuProIleSerLeuPheAspTyrPheIleValGlyGluGlyAsn 50  
 151 AGCTCCCAATCTCCCTATTGATATTATTATTGTTGGCGAGGAGGTAAT 200  
 51 GluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPheValLysLy 67  
 201 GAGGAGGAGGACACCTCACCTCCAGGGTTCGCTAATTTCTGAGAA 250  
 67 sGlnThrPheAsnLysValLysTrpTyrLeuGlyAlaArgCysHisIle 84  
 251 GCAAACTTTAATAAAGTAGTGAAGTGTATTGTTGGTGCGCGCTGCCACATCG 300  
 84 LuLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrCysSerLysGlu 100  
 301 AGAAAGCCAAAGGAAGTATGATCAGCAGAATAAAGAATATTGCAGTAAGAA 350  
 101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArgse 117  
 351 GGCACCTTACTTATTGAATGAGTGGAGCTCTCGATCTCAAGCAACAGGAG 400  
 117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuVal 134  
 401 TGACCTGTCTACTGCTGAGTACCTTTGTTGGAGCGGAGTCTGCTGA 450  
 134 hrValAlaGluGlnHisProValThrPheValLysAsnPheArgGlyLeu 150  
 451 CCCTTGCAAGACGACCCCTGTAAACGTTTGTCAAAAATTTCCCGCGGCTG 500  
 151 AlaGluLeuLeuLysValSerGlyLysMetGlnLysArgAspTrpLysTh 167

501 GCTGAACCTTTGAAGTGAGCGGAAATGCAAAAGCGTGATTGGAAC 550  
 167 rAsnValHisPheIleValGlyProGlyCysGlyLysSerLysTrpA 184  
 551 CAATGTACACTTCATTGTGGGGCCACCTGGGTGTGTAAGCAATGGG 600  
 184 laAlaAsnPheAlaAsnProGluThrThrTyrTrpLysProProLysAsn 200  
 601 CTGCTAATTTTGAACCCCGGAACACACATCTGGAACACCACTAAAAC 650  
 201 LysTrpTrpAspGlyTyrHisGlyGluLysValValValIleAspAspPh 217  
 651 AAGTGTGGGATGTTTACCATGTTGTAAGAAAGTGTGTTATTGATGACTT 700  
 217 eTyrGlyTrpLeuProTrpAspAspLeuLeuArgLeuCysAspArgTyrP 234  
 701 TTATGGCTGGCTCCGCGGGATGATCTACTGAGACTGTGTGATCGATC 750  
 234 roLeuThrValLysThrLysGlyGlyThrValProPheLeuAlaArgSer 250  
 751 CATTGACTGTAAACATAAAGTGAAGTGTACTCTTTTGGCCCGCAGT 800  
 251 IleLeuIleThrSerAsnGlnThrProLeuGluTrpTyrSerSerThrAl 267  
 801 ATTCTGATTACCAACATCAGACCCCGTTGGAATGTGTACTCTCACTGC 850  
 267 aValProAlaValGluAlaLeuTyrArgArgIleThrSerLeuValPhe 284  
 851 TGTCCCGAGCTGTAGAAGCTCTCTATCGGAGGATTACTTCTTGGTATTT 900  
 284 rPLysAsnAlaThrLysGlnSerThrGluGluGlyGlnPheValThr 300  
 901 GGAAGATGCTACAAACAATCCACGGAGGAGGAGGCGGCGGCTGCACC 950  
 301 LeuSerProCysProGluPheProTyrGluIleAsnTyr 314  
 951 CTTTCCCGCCCATGCCCTGAATTTCCATATGAATAAATATAC 992

seq\_name: gb\_vi:AF086836

seq\_documentation\_block:

LOCUS AF086836 1768 bp DNA circular VRL 29-SEP-1998  
 DEFINITION Porcine circovirus strain M226, complete genome.

ACCESSION AF086836

NID 93661521

VERSION AF086836.1 GI:3661521

KEYWORDS

porcine circovirus.

ORGANISM

Viruses; ssDNA viruses; Circoviridae; Circovirus.

REFERENCE 1 (bases 1 to 1768)

AUTHORS Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.

TITLE Emergence of a new porcine circovirus

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1768)

AUTHORS Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.

TITLE Direct Submission

JOURNAL Submitted (26-AUG-1998) VIDO, University of Saskatchewan, 120

Veterinary Road, Saskatoon, Saskatchewan S7N 5E3, Canada

FEATURES

Location/Qualifiers

1..1768

/organism="porcine circovirus"

/strain="M226"

/db\_xref="taxon:46221"

51..995

/codon\_start=1

/product="p35.8"

/protein\_id="AAC61741.1"

/db\_xref="PID:g3661522"

/db\_xref="GI:3661522"

/translation="WPSKNGRSQPQPHKRWVFTLNNPSEDERKKIRELPISLFDYFI  
 VGEENEGSRTPHLOGFANFYKQTFNKVKWFGARCHIEKAKGTDOONKEVCSKEGN

LLIECGAPRSQGRSDLSTAVSTLLESGILVTVAKHQHPVTVFNFRGLAEELKLVSGKM  
QKRDWNTNHFIVGPPCCGSKSWANFANPETTYWRPNKKNWDGHHGERKVVYIDDFY  
GWLPMDDLRLCDRYLFTVTKGTGTFPLARSILITSNOTPLEWYSTAVPAVEALYR  
RITSLVFWKNATKQSTEEGQFVTLSPPCPEFFPEINY  
complement(1034..1735)

## CDS

/codon\_start=1  
/product="P27.9"  
/protein\_id="AAC61742.1"  
/db\_xref="PID:g3661523"  
/db\_xref="GI:3661523"  
/translation="MYPRRYRRRRPRSHLGQILRRPWLPHRPHRYWRKNGI  
FNRLSKRTGTVTKRTVTTPSWADMRFKIDDFVPPGGTINKISIPFFYIRKVK  
VEWPCSPITQGRGVSFAVILDNFTKATYLDYVNYSSRHTIPQFFSYHSRY  
FTPKPVLDSITDYFOENNRNQLWLRLQTSNVDHVLGLSTAFENSKYDQYNIRVTMY  
VQPRENLKDPPLP"

BASE COUNT 463 a 363 c 480 g 462 t

## ORIGIN

alignment\_scores:  
Quality: 1710.00 Length: 314  
Ratio: 5.446 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 99.363

## alignment\_block:

US-09-209-961-3 x AF086836

Align seg 1/1 to: AF086836 from: 1 to: 1768

1 MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTr 17  
51 ATGCCCAAGCAAGCAATGGAAGAGCGGCCCAACCAACACATAAAGGTG 100  
17 pValPheThrLeuAsnAsnProSerGluAspGluArgLysLysLleArgG 34  
101 GGTGTTCACGCTCAATATCTTCGAGAGACGAGCGCAAGAAATACGGG 150  
34 LuLeuProIleSerLeuPheAspTyrPheLleValGlyGluGluGlyAsn 50  
151 AGCTCCCAATCTCCCTATTATTGATTATTTTGGTGGCAGAGGGGTAAAT 200  
51 GluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPheValLysLy 67  
201 GAGGAAGCCGCAACACCCACTCAGGGGTTCGCTAATTTGTGAAGAA 250  
67 sGlnThrPheAsnLysValLysTyrTrpLeuGlyAlaArgCysHisIleG 84  
251 GCAAACTTTTAATAAGTGAAGTGTATTTTGGTCCCGCTGCCACATCG 300  
84 LuLysAlaLysGlyThrAspGlnAsnLysGluTyrCysSerLysGlu 100  
301 AGAAAGCGAAAGGAAGTCAATCAGCAGATAAAGAAATATTCAGTAAGAA 350  
101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArgSe 117  
351 GCACACTTACTATTGATATGGAGCTCTCGATCTCAAGGACACACGGAG 400  
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyLleLeuValT 134  
401 TGACCTGTCTACTGCTGTGAGTACCTTCTTGAGAGCGGGATTCGTGTA 450  
134 hrValAlaGluGlnHisProValThrPheValLysAsnPheArgGlyLeu 150  
451 CCGTTGCAAGACGAGCAGCCCTGTAAGCTTTGTCAAAATTTCCGCGGGCTG 500  
151 AluGluLeuLeuLysValSerGlyLysMetGlnLysArgAspTrpLysTh 167  
501 GCTGAACATTTTGAAGAGTACGGGGAAATGCAAAACGCTGATTTGAAAC 550  
167 rAsnValHisPheLleValGlyProGlyCysGlyLysSerLysTrpA 184  
551 CAATGTACACTTCATTTGTTGGGCCACCTGGGTGTTGTTAAAGCAATGGG 600

184 laAlaAsnPheAlaAsnProGluThrThrTyrTrpLysProProLysAsn 200  
601 CTGCTAATTTTCAAAACCGGAACACCATCTACTGGAACACCACTAAAC 650  
201 LysTrpTrpAspGlyTyrHisGlyGluLysValValValLleAspAspPh 217  
651 AAGTGGTGGGATGGTTACCATGTTGAAAAAGTGGTTGTTATTGATGACT 700  
217 eTyrGlyTrpLeuProTrpAspAspLeuLeuArgLeuCysAspArgTyrP 234  
701 TTATGGCTGGCTGGCGTGGGATGATCTACTGACACTGTGTGATCGATATC 750  
234 roLeuThrValLysThrLysGlyGlyThrValProPheLeuAlaArgSer 250  
751 CATTGACTGTAAAACTAAAGTGGGAACTGTACCTTTTGGCCCGCAGT 800  
251 lleLeuIleThrSerAsnGlnThrProLeuGluTrpTyrSerSerThrAl 267  
801 ATTCTGATTACCAGCAATCAGACCCCGTGGATGTACTCTCTCAACTGC 850  
267 aValProAlaValGluAlaLeuTyrArgArgIleThrSerLeuValPheT 284  
851 TGTCCACGCTGTAGAAGCTCTCTATCGGAGGATTACTCTCTTGGTATTT 900  
284 rPLysAsnAlaThrLysGlnSerThrGluGluGlyGlnPheValThr 300  
901 GGAAGAATGCTACAAACAATCCACGGAGAGGGGCCAGTCGTGCACC 950  
301 LeuSerProProCysProGluPheProTyrGluIleAsnTyr 314  
951 CTTTCCCCCATGCTGCTGAATTTCCATATGAATAAATATAC 992

seq\_name: gb\_vi:AF086834

seq\_documentation\_block:

LOCUS AF086834 1768 bp DNA circular VRL 29-SEP-1998  
DEFINITION Porcine circovirus strain B9, complete genome.

ACCESSION AF086834

NID 93661515

VERSION AF086834.1 GI:3661515

KEYWORDS

SOURCE porcine circovirus.

ORGANISM porcine circovirus

REFERENCE 1 (bases 1 to 1768)

AUTHORS Wang,L., Willson,P., Chow,B., Gibbons,E. and Babiuk,L.

TITLE Emergence of a new porcine circovirus

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1768)

AUTHORS Wang,L., Willson,P., Chow,B., Gibbons,E. and Babiuk,L.

TITLE Direct Submission

JOURNAL Submitted (26-AUG-1998) VIDO, University of Saskatchewan, 120

Veterinary Road, Saskatoon, Saskatchewan S7N 5E3, Canada

FEATURES Location/Qualifiers

source

1..1768

/organism="porcine circovirus"

/strain="B9"

/db\_xref="taxon:46221"

51..995

/codon\_start=1

/product="P35.8"

/protein\_id="AAC61737.1"

/db\_xref="PID:g3661516"

/db\_xref="GI:3661516"

/translation="MPSKNGRSGPQPHKRWFTLNNPSEDERKKIRLPISLDFYI

VGEENEGEPTPHLOGFANFVKOTFNKVKYLGARCHIEKAKCTDOONKEYCSKEGN

LLIECGAPRSQGRSDLSTAVSTLLESGILVTVAKHQHPVTVFNFRGLAEELKLVSGKM

QKRDWNTNHFIVGPPCCGSKSWANFANPETTYWRPNKKNWDGHHGERKVVYIDDFY

GWLPMDDLRLCDRYLFTVTKGTGTFPLARSILITSNOTPLEWYSTAVPAVEALYR

RITSLVFWKNATKQSTEEGQFVTLSPPCPEFFPEINY

complement(1034..1735)

/codon\_start=1

/product="P27.9"

CDS

```
/protein_id="AAC61738.1"
/db_xref="pid:g3661517"
/db_xref="gi:3661517"
/translation="MTYPRRRYRRRRHRRSHLGOILRRPWLHVPHRRYRWRKNGI
ENTLSRFTGTVKATTTTTSWADVMRFNIDDFVPPGGTGKTKISIPFEYRIRKVK
VEFWPCSPITQDGRGVSTAVLDDNFVKATLYDPYVNYSSRHTIPQPFYSHSRY
FTPKPVLSTIDYFQPNKRNQLWLRLQTSRNVHVLGTAFENSKYDQDYNIRVTM
VQREFNLKDPKP"
BASE COUNT      463 a  357 c  482 g  466 t
ORIGIN
```

```
alignment_scores:
  Quality: 1706.00      Length: 314
  Ratio: 5.433          Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.045
```

```
alignment_block:
US-09-209-961-3 x AF086834
```

```
Align seg 1/1 to: AF086834 from: 1 to: 1768
```

```
1 MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTr 17
|||||
51 ATGCCCAAGCAAGAAATGGAAGAGCGGACCCCAACACACATAAAGGTG 100
|||||
17 pValPheThrLeuAsnAsnProSerGluAspGluArgLysLysLysLeArg 34
|||||
101 GGTGTTCCACCTGAATATCTCCGAAGAGCGGACGCAAGAAATACGGG 150
|||||
34 luLeuProIleSerLeuPheAspTyrPheIleValGlyGluGlyAsn 50
|||||
151 AGTCCCAATCTCCATTATGATTATTTATTTGTTGCGAGGAGGTAAT 200
|||||
51 GluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPheValLysLy 67
|||||
201 GAGGAGGAGCAACACCTCACCTCCAGGGTTCGTAATTTGTGAAGAA 250
|||||
67 sGlnThrPheAsnLysValLysTrpTyrLeuGlyAlaArgCysHisIleG 84
|||||
251 GCAAACTTTTAATAAGTGAAGTGGTATTTGGGTGGCCGCTGCCACATCG 300
|||||
84 luLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrCysSerLysGlu 100
|||||
301 AGAAGCCAAAGGAACGATCAGCAGATAAGAAATATTGTAGTAAGAA 350
|||||
101 GlyAsnLeuLeuLysGlyCysGlyAlaProArgSerGlnGlyGlnArgSe 117
|||||
351 GGCAACTTACTTATTGAATGTGGAGCTCTCGATCTCAAGGACACACGGAG 400
|||||
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValT 134
|||||
401 TGACCTGCTACTGCTGTGAGTACCTTTGTTGGAGAGCGGATCTGTGTA 450
|||||
134 hrValAlaGluGlnHisProValThrPheValLysAsnPheArgGlyLeu 150
|||||
451 CCGTTGCAAGACGACCCCTGTACAGTTTGTCAAAAATTTCCGGCGGCTG 500
|||||
151 AlaGluLeuLysValSerGlyLysMetGlnLysArgAspTrpLysTh 167
|||||
501 GCTCAACTTTTGAAGTGAGCGGAAAAATGCAAAAGCGTGATTGGAACAC 550
|||||
167 rAsnValHisPheIleValGlyProProGlyCysGlyLysSerLysTrpA 184
|||||
551 CAATGTACACTTATTGTGGGGCCACCTGGGTGTGTTAAAGCAAAATGGG 600
|||||
184 laAlaAsnPheAlaAsnProGluThrThrTyrTrpLysProProLysAsn 200
|||||
601 CTGCTAATTTTGAACCCGGAACACCATACTGGAACACCACTTAAACAC 650
|||||
201 LysTrpTrpAspGlyTyrHisGlyGluLysValValValLysAspAspPh 217
|||||
651 AAGTGGTGGGATGTTACCATGGTGAAGAAAGTGGTTGTTATTGATGACTT 700
```

```
seq_name: gb_vi:AF086835
```

```
seq_documentation_block:
```

```
LOCUS AF086835 1768 bp DNA circular VRL 29-SEP-1998
DEFINITION Porcine circovirus strain 9741, complete genome.
```

```
ACCESSION AF086835
```

```
NID g3661518
```

```
VERSION AF086835.1 GI:3661518
```

```
KEYWORDS
```

```
SOURCE porcine circovirus.
```

```
ORGANISM porcine circovirus
```

```
REFERENCE 1 (bases 1 to 1768)
```

```
AUTHORS Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.
```

```
TITLE Emergence of a new porcine circovirus
```

```
JOURNAL Unpublished
```

```
REFERENCE 2 (bases 1 to 1768)
```

```
AUTHORS Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.
```

```
TITLE Direct Submission
```

```
JOURNAL Submitted (26-AUG-1998) VIDO, University of Saskatchewan, 120
```

```
Veterinary Road, Saskatoon, Saskatchewan S7N 5E3, Canada
```

```
FEATURES
```

```
source
```

```
1..1768
```

```
/organism="porcine circovirus"
```

```
/strain="9741"
```

```
/db_xref="taxon:46221"
```

```
51..995
```

```
/codon_start=1
```

```
/product="p35.8"
```

```
/protein_id="AAC61739.1"
```

```
/db_xref="PID:g3661519"
```

```
/db_xref="GI:3661519"
```

```
/translation="MPSKNGSGGPGPHRWVFTLNPSNKKKIRELPISLDFPI
VGEENEGRTPLQGFANFKQTFNKVKYLGAHCHEKAGTDOONKEYCSKEGN
LLIEGAPRSQGSQSDLSSTAVSTLLESGLVTVAKQHPVTFVKNFRGLAEKLVSGKM
QKRDKTNVHFIVGPPGCKSKWAFANPETTYWPKPKNWDGYHGEKVVIDDFY
GMLPMDLLRLCDRYPLTVTKGTVPPLARSLITSNQTPLEWYSSSTAVPAVEALYR
RTSLVFNKATEQSTEGGQFVTLSPCPPEPPEYIN"
```

```
complement(1034..1735)
```

```
/codon_start=1
```

```
/product="p27.9"
```

```
/protein_id="AAC61740.1"
```

```
/db_xref="PID:g3661520"
```

```
/db_xref="GI:3661520"
```

```
/translation="MTYPRRRYRRRRHRRSHLGOILRRPWLHVPHRRYRWRKNGI
ENTLSRFTGTVKATTTTTSWADVMRFNIDDFVPPGGTGKTKISIPFEYRIRKVK
VEFWPCSPITQDGRGVSTAVLDDNFVKATLYDPYVNYSSRHTIPQPFYSHSRY
FTPKPVLSTIDYFQPNKRNQLWLRLQTSRNVHVLGTAFENSKYDQDYNIRVTM"
```

BASE COUNT 464 a 360 c 480 g 464 t  
ORIGIN  
VOFRENLKDPPLKP\*  
alignment\_scores:  
Quality: 1701.00 Length: 314  
Ratio: 5.417 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 98.726  
alignment\_block:  
US-09-209-961-3 x AF086835 ..  
Align seg 1/1 to: AF086835 from: 1 to: 1768  
1 MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTr 17  
51 ATGCCCAAGCAAGAGAAATGGAAGAGCGGAGCCCAACACCAATAAAGGTG 100  
17 pValPheThrLeuAsnAsnProSerGluAspGluArgLysLysIleArg 34  
101 GGTGTTCACGCTGATATCTCTCCGAACCAAGCGCAAGAAATACGG 150  
34 LuLeuProIleSerLeuPheAspTyrPheIleValGlyGluGlyAsn 50  
151 AGCTCCCAATCTCCCTATTGATTATTTATTGTTGGCGAGGAGGTAAT 200  
51 GluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPheValLysLy 67  
201 GAGGAAGACCAACACCTCACCTCCAGGGGTTCGCTAATTTTGTGAAGA 250  
67 sGlnThrPheAsnLysValLysTyrTyrLeuGlyAlaArgCysHisIleG 84  
251 CCAACCTTTTAAAGTGAAGTGTATTTGGTGGCCCGTCCACATCG 300  
84 LuLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrCysSerLysGlu 100  
301 AGAAGCCAAAGGAACGTATCAGCAAAATAAAGATATTCAGTAAAGAA 350  
101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArgSe 117  
351 GGCAACTTTTACTATTGATGTGGAGCTCTCGATCTCAAGGACACGCGAG 400  
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValT 134  
401 TGACCTGTCTACTGTGTGAGTACTTGTGGAGAGCGGATTCGTGGA 450  
134 hrValAlaGluGlnHisProValThrPheValLysAsnPheArgGlyLeu 150  
451 CCGTGTCAACACGACACCTCTACGTTTGTCAAAATTTCCGGGGGTG 500  
151 AlaGluLeuLysValSerGlyLysMetGlnLysArgAspTyrLysTh 167  
501 GCTGAACCTTTTGAAGTGAAGGGAATGCAAAAGCGTGTGGAAGAAC 550  
167 rAsnValHisPheIleValGlyProGlyCysGlyLysSerLysTrpA 184  
551 CAATGTACACTTATTGTGGGGCCACCTGGTGTGTTAAAGCAAAATGGG 600  
184 laAlaAsnPheAlaAsnProGluThrThrTyrTyrLysProProLysAsn 200  
601 CTGCTAATTTTGAACCCGGAACCACTACTGGAACCAACCACTAAAGAC 650  
201 LysTrpTrpAspGlyTyrHisGlyGluLysValValIleAspAspPh 217  
651 AAGTGTGGGATGTACCATGGTGAAGAAAGCTGTTGTATGATGACTT 700  
217 eTyrGlyTrpLeuProTrpAspLeuLeuArgLeuCysAspArgTyrP 234  
701 TTATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 750  
234 roLeuThrValLysThrLysGlyThrValProPheLeuAlaArgSer 250

751 CATTGACTGTAAAACTAAAGGTGGAACGTGTACCTTTTGGCCCGCAGT 800  
251 ileLeuIleThrSerAsnGlnThrProLeuGluTyrTyrSerSerThrAl 267  
801 ATCTGATTACCAAGCAATCAGACCCCGTTGGAAAGGTACTCTCTCAACTGC 850  
267 aValProAlaValGluAlaLeuTyrArgArgIleThrSerLeuValPhe 284  
851 TGTCCACGCTGTAGAAGCTCTCTATCGGAGGATTACTCTCTGTGATTTT 900  
284 rPlysAsnAlaThrLysGlnSerThrGluGluGlyGlyGlnPheValThr 300  
901 GGAAGAATGCTACAGAACAATCCACGAGGAGGGGCCAGTTTGTCAACC 950  
301 LeuSerProProCysProGluPheProTyrGluIleAsnTyr 314  
951 CTTTCCCCCATGCCCTGAATTTCATATGAATAAATATAC 992  
seq\_name: gb\_v1:AF055391  
seq\_documentation\_block:  
LOCUS AF055391 1768 bp DNA circular VRL 13-SEP-1998  
DEFINITION Porcine circovirus Type II from USA, complete genome.  
ACCESSION AF055391  
NID G3598796  
VERSION AF055391.1 GI:3598796  
KEYWORDS  
SOURCE porcine circovirus.  
ORGANISM porcine circovirus  
Viruses; ssDNA viruses; Circoviridae; Circovirus.  
REFERENCE  
1 (bases 1 to 1768)  
AUTHORS Meehan,B.M., McNeilly,F., Todd,D., Kennedy,S., Jewhurst,V.A.,  
Ellis,J.A., Hassard,L.E., Clark,E.G., Haines,D.M. and Allan,G.M.  
TITLE Characterization of novel circovirus DNAs associated with wasting  
syndromes in pigs  
JOURNAL J. Gen. Virol. 79 (pt 9), 2171-2179 (1998)  
MEDLINE 98418498  
REFERENCE 2 (bases 1 to 1768)  
AUTHORS Meehan,B.M.  
TITLE Direct Submission  
JOURNAL Submitted (26-MAR-1998) Brian M Meehan, Veterinary Science, The  
Queen's University of, Belfast, Stoney Road, Belfast, Northern  
Ireland, BT4 3SD, U.K.  
FEATURES  
source  
1. .1768  
/organism="porcine circovirus"  
/specific\_host="Sus scrofa"  
/strain="Porcine circovirus Type II"  
/db\_xref="taxon:46221"  
/clone="Imp.999 Ecori No.8"  
/country="USA"  
/complement(101..190)  
/codon\_start=1  
/product="ORF6"  
/protein\_id="AAC35308.1"  
/db\_xref="PID:g3598807"  
/db\_xref="GI:3598807"  
/translation="MASSTPASPAPSDILSRPQSERPPGRWT"  
103..210  
/codon\_start=1  
/product="ORF10"  
/protein\_id="AAC35306.1"  
/db\_xref="PID:g3598805"  
/db\_xref="GI:3598805"  
/translation="MSTAQEGVLTVALTVYVKVRRRLKMPFFLLQR"  
261..320  
/codon\_start=1  
/product="ORF7"  
/protein\_id="AAC35307.1"  
/db\_xref="PID:g3598806"  
/db\_xref="GI:3598806"  
/translation="MAAGVSSSNVTPPWIRHS"  
complement(311..439)  
CDS  
CDS  
CDS

```
/note="predicted 4.6 kDa protein"
/codon_start=1
/product="ORF9"
/protein_id="AAC35304.1"
/db_xref="PID:g3598803"
/db_xref="GI:3598803"
/translation="MWLGASASSILLAGHVAEEVLPCCRCRSALVILTAHFFRFQL"
398..1342
/codon_start=1
/product="putative Rep protein (ORF1)"
/protein_id="AAC35298.1"
/db_xref="PID:g3598797"
/db_xref="GI:3598797"
/translation="MPSKKNRSGPQPKRWFTLNNPSEDERKKIRPLISLDFYFI
VGEENEGSRPHLGAFNFVKQTFNKVYLGARCHEKAKGTDQONKEVCSEKGN
LLIECGAPSGQSRSDLSSTAVSLLESGLTVVAQHPVTFVRNFRGLAELLKYSKGM
QKRDKTNHVIIVGPPGSKAAANFADPETTYWKPPRNKWDGTHGEEVVVIDFY
GWLPMDDLRLCDRYPLVETGKGVFPLARSILITSNOTPLEWYSYTAAPAVEALYR
RITSLVFWKNATEOSTEGGQFVTLSPPCPEFPYBINT"
674..679
complement(704..1018)
/note="predicted 11.9 kDa protein"
/codon_start=1
/product="ORF3"
/protein_id="AAC35300.1"
/db_xref="PID:g3598799"
/db_xref="GI:3598799"
/translation="MTVITPLVSRWFPVCGFRVCKIISPFATFTPRPHNDVYIGLPI
TLLHFPFHQFQSOAPEISDKRYRVLCCNGHGTALQOQTHSSROVTPLSLRSRSTF
NK"
complement(733..912)
/note="predicted 6.5 kDa protein"
/codon_start=1
/product="ORF4"
/protein_id="AAC35301.1"
/db_xref="PID:g3598800"
/db_xref="GI:3598800"
/translation="MTCILVFQSRFCIFPLTFKSSASPRKFLTNVTGCCSATVTRPL
SNKVLTAVDRLRGP"
complement(1035..1100)
/note="predicted 2.3 kDa protein"
/codon_start=1
/product="ORF8"
/protein_id="AAC35303.1"
/db_xref="PID:g3598802"
/db_xref="GI:3598802"
/translation="MDIDHTVSDHPTAASHKSHQ"
1330..1335
complement(1336..1380)
/note="predicted 1.8 kDa protein"
/codon_start=1
/product="ORF11"
/protein_id="AAC35305.1"
/db_xref="PID:g3598804"
/db_xref="GI:3598804"
/translation="MNNKNHYEVIKTKQ"
1363..1524
/note="predicted 6.2 kDa protein"
/codon_start=1
/product="ORF5"
/protein_id="AAC35302.1"
/db_xref="PID:g3598801"
/db_xref="GI:3598801"
/translation="MVFIIHLGFKWGVFKIFSELYIHGYTDIVLVLVVTFERSAEA
YVWHISRGL"
complement(1369..1374)
complement(join(1381..1768,1..314))
/note="predicted 27.8 kDa protein"
/codon_start=1
/product="ORF2"
/protein_id="AAC35299.1"
/db_xref="PID:g3598798"
/db_xref="GI:3598798"
```

```
/translation="MTYPRRRYRRRRHPRSHLQILRRRPLVHPRHRYRWRKNGI
FNTRLSRFGYTVKATVTPSWAVMMNFIDDFVPGGYNKISIPPEYRIKKVK
VEFWPCSPITQDGRGVSTAVILDNDNFVKATALTVDYVYNTSSRHTIPQPFYHSRY
FTPKPVLSDTIDYFQPNKRNQIWLQTSRNVHDVGLGTAFENSIDQDYNIRVTM
VOFREFLNKDPLPKP"
BASE COUNT 453 a 360 c 494 g 461 t
ORIGIN

alignment_scores:
  Quality: 1682.00      Length: 314
  Ratio: 5.391         Gaps: 0
  Percent Similarity: 99.363  Percent Identity: 97.452

alignment_block:
US-09-209-961-3 x AF055391
Align seg 1/1 to: AF055391 from: 1 to: 1768

1 MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTr 17
|||||
398 ATGCCAGCAGAGAAGATGGAAGAGCGGACCCCAACCAACATATAAGGTG 447

17 pValPheThrLeuAsnAsnProSerGluAspGluArgLysLysIleArg 34
|||||
448 GGTGTTCCAGCGTGAATATCTCCGAAGACGAGCGCAAGAAATACGGG 497

34 LuLeuProIleSerLeuPheAspTyrPheIleValGlyGluGlyAsn 50
|||||
498 AGCTCCCAATCTCCTATTGATTATTATTATTGTCGAGGAGGGTAAAT 547

51 GluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPheValLys 67
|||||
548 GAGGAAGACGACACCTCACCTCAGGGGTTCGCTAATTTTGTGAAGAA 597

67 sGlnThrPheAsnLysValLysTyrTyrLeuGlyAlaArgCysHisIle 84
|||||
598 GCAAACTTTTAATAAGTGAAGTGTATTGGTGGCGCGCTGCCACATCG 647

84 LuLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrCysSerLysGlu 100
|||||
648 AGAAAGCCAAAGGAACATGATCAGCAGATAAAGAAATATTGCAGTAAAGAA 697

101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArgSe 117
|||||
698 GGCACCTTACTATTGATGTGAGGAGCTCTCGATCTCAAGACACACGGAG 747

117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuVal 134
|||||
748 TGACCTGTCTACTGCTGTGACTACCTTCTTGGAGAGCGGAGTCTGGTGA 797

134 hrValAlaGluGlnHisProValThrPheValLysAsnPheArgGlyLeu 150
|||||
798 CCGTGTGAGAGCAGCAGCCCTGTAACTGTTGTCAAGAAATTTCCCGCGGCTG 847

151 AlaGluLeuLeuLysValSerGlyLysMetGlnLysArgAspTrpLysTh 167
|||||
848 GCTGAACCTTTTGAAGTGAGCGGAAAATGCAAGACGCTGATTGGAAGAC 897

167 rAsnValHisPheIleValGlyProGlyProGlyLysLysSerLysTrpA 184
|||||
898 CAATGTACACGCTCATTTGGGGGCCACCTGGGTGTGTTAAAGCAATAGG 947

184 laAlaAsnPheAlaAsnProGluThrThrTyrTrpLysProProLysAsn 200
|||||
948 CTGCTAATTTTGCAGACCCGGAACACCATACTTGGAAACACCACTAGAAC 997

201 LysTrpTrpAspGlyTyrHisGlyGluLysValValValIleAspAspPh 217
|||||
998 AAGTGGTGGGATGGTTACCATGGTGAAGAAAGTGGTTGTTATTGATGACT 1047

217 eTyrGlyTrpLeuProTrpAspAspLeuLeuArgLeuCysAspArgTyrP 234
|||||
```





```

/product="ORF2"
/protein_id="AAC35310.1"
/db_xref="PID:g3598810"
/db_xref="GI:3598810"
/translation="MTYPRRRYRRRRPRLSHGLQLRRPWLVPHPHRYRWRKNGI
FNRLSRFTGYTKRTVTTPSWAVDMRFKIDFVPGGNTKISIPFEYIRKVK
VFVPCSPITQDGRGVSTAVILDDNFVKATALYDVPVNTSSRHTIPQPFYSHSRY
FTPKPVLDSIDYFQPNRRNLWLRLQTSNVDHVLGLAAFENSKYDQDYNIRVTMY
VQFRENLADPLKP"
BASE COUNT 451 a 361 c 495 g 461 t
ORIGIN

```

```

alignment_scores:
  Quality: 1682.00      Length: 314
  Ratio: 5.391          Gaps: 0
  Percent Similarity: 99.363  Percent Identity: 97.452

alignment_block:
US-09-209-961-3 x AF055392 ..
Align seg 1/1 to: AF055392 from: 1 to: 1768

```

```

1 MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTr 17
|||||
398 ATGCCAGCAAGAAGATGGAAGAGCGGAGCCCAACACACATAAAGGTG 447
17 pValPheThrLeuAsnAsnProSerGluAspGluArgLysLysIleArgG 34
|||||
448 GGTGTTCCAGCTGAATAATCTTCCGAAGACGAGCGCAAGAAATACGGG 497
34 LuLeuProIleSerLeuPheAspTyrPheIleValGlyGluGlyAsn 50
|||||
498 AGCTCCCAATCTCCTATTGATTAATTTATTTGTCGCGAGAGGATTAAT 547
51 GluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPheValLysLy 67
|||||
548 GAGGAAGACGACACACCTCACCTCCAGGGTTCGTAATTTTGTGAAGA 597
67 sGlnThrPheAsnLysValLysTyrPyrLeuGlyAlaArgCysHisIleG 84
|||||
598 GCAACCTTTTATAAAGTGAAGTGTATTGGTGGTCCGCTGCCACATCG 647
84 LuLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrCysSerLysGlu 100
|||||
648 AGAAGCCAAAGGAACCTGATCAGCAGATAAAGATAATGTCAGATAAGAA 697
101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArgSe 117
|||||
698 GGCAACTTACTTATTGATGTGGAGCTCTCGATCTCAAGGACACGGAG 747
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValT 134
|||||
748 TGACCTGTCTACTGCTGTGAGTACCTTGTGGAGCGGGAGTCTGGTGA 797
134 hrValAlaGluGlnHisProValThrPheValLysAsnPheArgGlyLeu 150
|||||
798 CCGTGGCAGACGACGCCCTGTACGTTTTCAGAAATTCGCGGGGCTG 847
151 AlaGluLeuLeuLysValSerGlyLysMetGlnLysArgAspTyrLysTh 167
|||||
848 GCTGAACCTTTGAAGTGGCGGGAATGAGAGCGTGTGATTGGAAGAC 897
167 rAsnValHisPheIleValGlyProGlnGlyCysGlyLysSerLysTrpA 184
|||||
898 CAATGTACACCTCATTTGTGGGCCACCTGGGTGTGTTAAACCAATGGG 947
184 laAlaAsnPheAlaAsnProGlnThrThrTyrTrpLysProProLysAsn 200
|||||
948 CTGCTAATTTTTCAGACCCGGAACACACATCTCGAAACACCTAGAAAC 997
201 LysTrpTrpAspGlyTyrHisGlyGluLysValValIleAspAspPh 217
|||||

```

```

998 AAGTGGTGGGATGTTACCATGGTCAAGAGCTGTTGTTATTGATGACTT 1047
217 eTyrGlyTrpLeuProTrpAspAspLeuLeuArgLeuCysAspArgTyrP 234
|||||
1048 TTATGGCTGGCTGGCGGGGATGATCTACTGAGACTGTGTGATCGATATC 1097
234 roLeuThrValLysThrLysGlyGlyThrValProPheLeuAlaArgSer 250
|||||
1098 CATTGACTGTAGACATAAAGGTGGAACCTGTACCTTTTGGCCCGCAGT 1147
251 IleLeuIleThrSerAsnGlnThrProLeuGluTrpTyrSerSerThrAl 267
|||||
1148 ATTCTGATTACCAACAATCAGACCCCGTTGAATGGTACTCCTCAACTGC 1197
267 aValProAlaValGluAlaLeuTyrArgArgIleThrSerLeuValPheT 284
|||||
1198 TGTCCAGCTGTAGAAGCTCTCTATCGAGGATTACTTCTTGGTATTTT 1247
284 rPlysAsnAlaThrLysGlnSerThrGluGluGlyGlyGlnPheValThr 300
|||||
1248 GGAAGATGCTACAGACAATCCACGGAGGAGGGGCCAGTTCGTCAACC 1297
301 LeuSerProProCysProGluPheProTyrGluIleAsnTyr 314
1298 CTTTCCCGCCCATGCCCTGAATTTCCATATGAATAAATTAC 1339

```

seq\_name: gb\_vi:AF109397

seq\_documentation\_block:

LOCUS AF109397 1768 bp DNA circular VRL 02-FEB-1999  
DEFINITION Bovine circovirus, complete genome.

ACCESSION AF109397

NID 94210313

VERSION AF109397.1 GI:4210313

KEYWORDS

SOURCE bovine circovirus.

ORGANISM bovine circovirus

Viruses: ssDNA viruses; Circoviridae; Circovirus.

REFERENCE 1 (bases 1 to 1768)

AUTHORS Hamel,A.I. and Nayar,G.P.S.

TITLE Nucleotide sequence of a circovirus detected in cattle with various clinical syndromes

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1768)

AUTHORS Hamel,A.I. and Nayar,G.P.S.

TITLE Direct Submission

JOURNAL Submitted (27-NOV-1998) Virology Laboratory, Veterinary Diagnostic Laboratory, Manitoba Agriculture, 545 University Crescent, Winnipeg, Manitoba R3T 5S6, Canada

REFERENCE 3 (bases 1 to 1768)

AUTHORS Hamel,A.I. and Nayar,G.P.S.

TITLE Direct Submission

JOURNAL Submitted (02-FEB-1999) Virology Laboratory, Veterinary Diagnostic Laboratory, Manitoba Agriculture, 545 University Crescent, Winnipeg, Manitoba R3T 5S6, Canada

REMARK Sequence update by submitter

COMMENT On Feb 2, 1999 this sequence version replaced gi:4028609.

FEATURES Location/Qualifiers

source

1..1768

/organism="Bovine circovirus"

/viral

/db\_xref="taxon:85542"

/note="BCV

sequence was obtained from several overlapping PCRs using DNA extracted from various tissues from cattle; similar to Porcine circovirus sequence presented in GenBank Accession Number AF027217"

repeat\_region 13..35

/rpt\_type-tandem

/rpt\_unit-13..18

51..995

CDS

/note="ORF-1; similar to Porcine circovirus ORF1 encoded by the sequence presented in GenBank Accession Number

```
AF027217.
/codon_start=1
/product="putative Rep and coat protein"
/protein_id="AAD11928.1"
/db_xref="PID:g4210314"
/db_xref="GI:4210314"
/translation="MPSKNGSGPOPHKRWVFTLNPSDEKRIKIPISLFDYFI
VEEGNEGRTHLOGFANFVKOTENKYKWLIGARCHIEKAGTDOONKEYCKEKN
LIEGAPRSQSORSLTAVSLLESGLVTVAEQHPVTFVFRFGLAELLKVSKEK
QRDKTNVHVIVPGCCCKSKWAANFADPETIYAPPRKNKWDGSHGEVVIDFY
GMLPMDLLRCDRLPTVETGKGTFTVLARSILITSNQTPLEWYSTAVPAVEALYR
RITSLVFNKATEQSTEEGGQFVTLSPPCPEFFIYNY"
polyA_signal
CDS
    complement(357..671)
    /note="similar to Porcine circovirus ORF-3 encoded by the
    sequence presented in GenBank Accession Number AF027217."
    /codon_start=1
    /product="ORF-3"
    /protein_id="AAD11930.1"
    /db_xref="PID:g4210315"
    /db_xref="GI:4210315"
    /translation="MTIPLVSRWFPVCGFRVCKISSPPAFITPRWPHNDVIGLPI
    TLLHFAHQKFSQPAEISDKRYRVLCCNGHOTPALQQGTHSSRQVTPLSLRSSTF
    NK"
CDS
    complement(386..565)
    /note="similar to Porcine circovirus ORF-4 encoded by the
    sequence presented in GenBank Accession Number AF027217."
    /codon_start=1
    /product="ORF-4"
    /protein_id="AAD11931.1"
    /db_xref="PID:g4210316"
    /db_xref="GI:4210316"
    /translation="MTCTLVFSRFCIFPLTFKSSASPRKFLINVTGCCSATVTRLPL
    SNKLTAVDRSLRCP"
CDS
    553..732
    /codon_start=1
    /product="ORF-12"
    /protein_id="AAD11927.1"
    /db_xref="PID:g4028610"
    /db_xref="GI:4028610"
    /translation="MTSLWHLGVVYKANGLLILOTRKKPTGNHLETSGGMVTWVKW
    LLLMTFMAGCRGIY"
CDS
    complement(688..753)
    /codon_start=1
    /product="ORF-8"
    /protein_id="AAD11935.1"
    /db_xref="PID:g4028618"
    /db_xref="GI:4028618"
    /translation="MDIDHTVSDVHDPTAASHKSHQ"
polyA_signal
CDS
    983..988
    complement(989..1033)
    /codon_start=1
    /product="ORF-11"
    /protein_id="AAD11937.1"
    /db_xref="PID:g4028620"
    /db_xref="GI:4028620"
    /translation="MNNKNHYEVIKTKQ"
    1016..1177
    /codon_start=1
    /product="ORF-5"
    /protein_id="AAD11932.1"
    /db_xref="PID:g4028615"
    /db_xref="GI:4028615"
    /translation="WVFIIHLGFKWGVKIKFSELYIHGYTDIWLVLWVTFVRSAAE
    YVHISRGF"
polyA_signal
CDS
    complement(1022..1027)
    complement(1256..1735)
    /note="similar to Porcine circovirus ORF-2 encoded by the
    sequence presented in GenBank Accession Number AF027217."
    /codon_start=1
    /product="ORF-2"
    /protein_id="AAD11929.1"
    /db_xref="PID:g4028612"

/db_xref="GI:4028612"
/translation="MTYPRRRRRRTPRSHLGOILRRPWLHPRHRYRRWRKNGI
FNTRLSRTFGYTVKATVTRPSWADMMRFNIDFVPPGCTNKISIPFEYRIKVK
VEFWPCSPITQDGRGVGSTAVILDNDNFVTKATALTIDPYVNYSSRHTIPQFYSYHSR"
CDS
    complement(1522..1611)
    /codon_start=1
    /product="ORF-6"
    /protein_id="AAD11933.1"
    /db_xref="PID:g4028616"
    /db_xref="GI:4028616"
    /translation="MASSTPASPPASDILSRLOSERPPGRWT"
    1524..1631
    /codon_start=1
    /product="ORF-10"
    /protein_id="AAD11936.1"
    /db_xref="PID:g4028619"
    /db_xref="GI:4028619"
    /translation="MSTAQEGVLTVALTVYPKVRERVLKMPFFLLQR"
    1682..1741
    /codon_start=1
    /product="ORF-7"
    /protein_id="AAD11934.1"
    /db_xref="PID:g4028617"
    /db_xref="GI:4028617"
    /translation="MAAGAGPSSAVTPPWIRHS"
    join(1750..1768,1..13)
    join(1762..1768,1..12)
    /note="putative; similar to the nonanucleotide motif of
    Porcine circovirus presented in GenBank Accession Numbers
    AF027217 and U49186"
BASE COUNT 452 a 362 c 493 g 461 t
ORIGIN

alignment_scores:
    Quality: 1682.00      Length: 314
    Ratio: 5.391          Gaps: 0
    Percent Similarity: 99.363      Percent Identity: 97.452

alignment_block:
    US-09-209-961-3 x AF109397 ..
    Align seg 1/1 to: AF109397 from: 1 to: 1768

1 MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTr 17
|||||
51 ATGCCAGCAAGAAGATGGAAGAGCGGACCCACCAACACATATAAGGTG 100
|||||
17 pValPheThrLeuAsnAsnProSerGluAspGluArgLysLysIleArg 34
|||||
101 GGTGTTACGCTGAATAATCTTCGAGAGCAGCGCAAGAAATACGGG 150
|||||
34 LuLeuProIleSerLeuPheAspTyrPheIleValGlyGluGluGlyAsn 50
|||||
151 ACCTCCCAATCCCTATTATTGATTATTTATTGTCGCGAGGAGGTAAT 200
|||||
51 GluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPheValLysLy 67
|||||
201 GAGGAAGGACGAACACCTCACCTCCAGGGGTCGCTAATTTCTGAAGAA 250
|||||
67 sGlnThrPheAsnLysValLysTyrTrpTyrLeuGlyAlaArgCysHisIle 84
|||||
251 GCAAACTTTTAATAAAGTGAAGTGTATTGTTGGTCCCGCTGCCACATCG 300
|||||
84 LuLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrCysSerLysGlu 100
|||||
301 AGAAAGCCCAAGAACTGATGACAGAATAAAGAAATATTGCAGTAAAGAA 350
|||||
101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArgSe 117
|||||
351 GCGCACTTACTTATTGAATGTGGAGCTCCTCGATCTCAAGGACACGGAG 400
|||||
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuVal 134
```

```
|||||
401 TGACCTGCTCTACTGCTGAGTACCTTTGTTGGAGCGGAGTCTGCTGA 450
134 hrValaGluGlnHisProValThrPheValLysAsnPheArgGlyLeu 150
451 CCCTTCAGACGACGACCCCTGTACACGTTTGTGAGAAATTTCCGCGGCTG 500
151 AlaGluLeuLysValSerGlyLysMetGlnLysArgAspTrpLysTh 167
501 GCTGACCTTTGAAAGTAGCGGGAATAATGCAGAGCGTGAATGGAAGAC 550
167 rAsnValHisPheIleValGlyProProGlyCysGlyLysSerLysTrpA 184
551 CAATGTACACGTCATTTGCGGGCCACCTGGGTGTGTAAAGCAATGGG 600
184 laalaAsnPheAlaAsnProGluThrThrTrpLysProLysAsn 200
601 CTGCTAATTTTGAGACCCCGGAACCATACACTGGAACACCATAGAAAC 650
201 LysTrpTrpAspGlyTyrHisGlyGluLysValValLysValLysAspPh 217
651 AAGTGGTGGATGGTACCATGGTGAAGAGTGGTGTATTGATGACTT 700
217 eTyrGlyTrpLeuProTrpAspPheLeuArgLysCysAspArgTyrP 234
701 TTATGGTGGCTGCGGTGGATGATCTACTGAGACTGTGTGATGATATC 750
234 roLeuThrValLysThrLysGlyGlyThrValProPheLeuAlaArgSer 250
751 CATGTAGTGTAGACATAAAGTGGAGTACTTACTTTTGGCCCGCAGT 800
251 IleLeuThrSerAsnGlnThrProLeuGluTrpTyrSerSerThrAl 267
801 ATTCTGATTACAGCAATACAGACCCGTTGGAATGGTACTCTCTCACTGC 850
267 aValProAlaValGluAlaLeuTyrArgArgIleThrSerLeuValPhe 284
851 TGTCCCGAGCTGTAGAAGCTCTATCGGAGGATTACTTCTCTGTTATTT 900
284 rPlysAsnAlaThrLysGlnSerThrGluGluGlyGlnPheValThr 300
901 GGAAGAATGTACAGAACAAATCACGGAGGAAGGGGCCAGTTCGTCACC 950
301 LeuSerProProCysProGluPheProTyrGluIleAsnTyr 314
951 CTTTCCCCCAGTCCCTGAAATTCATATGAATAAATATAC 992
```

seq\_name: gb\_vi:AF147751

```
seq_documentation_block:
LOCUS AF147751 1768 bp DNA VRL 08-JUN-1999
DEFINITION Porcine circovirus type 2 unknown genes.
ACCESSION AF147751
NID 95007010
VERSION AF147751.1 GI:5007010
KEYWORDS
SOURCE
  porcine circovirus type 2.
  porcine circovirus type 2.
REFERENCE
  1 (bases 1 to 1768)
  Poganichnyy,R., Yoon,K.-J., Harms,P., Swenson,S., Zimmerman,J. and
  Sorden,S.
  Characterization of clinical and immune responses in young swine to
  experimental porcine circovirus type II infection
JOURNAL
  Unpublished
REFERENCE
  2 (bases 1 to 1768)
  Yoon,K.-J. and Poganichnyy,R.M.
  Direct Submission
JOURNAL
  Submitted (30-APR-1999) Veterinary Diagnostic Laboratory, Iowa
  State University, College of Veterinary Medicine, Ames, Iowa 50011,
  USA
FEATURES
  source 1..1768
```

```
/organism="porcine circovirus type 2"
/isolate="ISUVDL 98-15237"
/specific_host="Sus scrofa"
/db_xref="taxon:85708"
/notes="isolated from pigs with postweaning multisystemic
wasting syndrome"
51..995
/notes="ORF1: 35-36kd"
/codon_start=1
/product="unknown"
/protein_id="AAD37776.1"
/db_xref="PID:g5007011"
/db_xref="GI:5007011"
/translation="MPSKNGRSGPOPHKRWFTLNPSDERKKIRELPISLDFYFI
VGEEGNEGRTHLQGFANFKQTKNKVKGWLGARCHIIEKAGTDOQNEKVCSEGN
LLIECAPRSQGRSDLSAVSTLSEGSlyTVAEQHPVTFVFRGLAEALLKVSQGM
QRKDKINHVIVGPGCGSKWAANFADPTTYWPKRKNWWDGYHGEVYVLDIFY
GWLWDLLRLCDRQPLTVEYTKGTFVFLARSILITSNQTPLEWYSTAVPAVEALYR
RITSLVFNKATEQSTEGGQFVTLSPPCPEFFEYNI"
complement(1034..1735)
/notes="ORF2: 27-28kd"
/codon_start=1
/product="unknown"
/protein_id="AAD37777.1"
/db_xref="PID:g5007012"
/db_xref="GI:5007012"
/translation="MTYPRRRYRRRRPRLSHLGOILRRPMLVHPRRYWRBKNGI
FNRLSRTGYTKATVTPTSPWADMRENIDDFVPPGGGKNKISIPFYRIRKVK
VEFWPCSPITQDGRGVSTAVILDNFVTKATATLDYVNYISRRHTIPQFSCHSRY
FTPKPFLDSTIDYFQPNKRNQLWLRLQTSRNVHDVHGLGTAFAENSIYDQDNIKRTMY
VOFRENLKDPPLKP"
BASE COUNT 450 a 362 c 494 g 462 t
ORIGIN

alignment_scores:
  Quality: 1682.00      Length: 314
  Ratio: 5.391         Gaps: 0
  Percent Similarity: 99.363  Percent Identity: 97.452

alignment_block:
US-09-209-961-3 x AF147751 ..
Align seg 1/1 to: AF147751 from: 1 to: 1768
1 MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTr 17
51 ATGCCCAAGAAAGAAATGGAAGAGCGGACCCCAACACACATAAAAGGTG 100
17 pValPheThrLeuAsnAsnProSerGluAspGluArgLysLysIleArgG 34
101 GGTGTTACGCTGAATAATCTCTCGAAGACGAGCGCAAGAAATACGGG 150
34 LuLeuProIleSerLeuPheAspTyrPheIleValGlyGluGlyAsn 50
151 AGTCCCAATCTCCCTATTGATTATTTATTGTTGGCAGGAGGTAAT 200
51 GluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPheValLysLy 67
201 GAGGAAGACGACCAACCTCCAGGAGTTCGCTAATTTTGTGAAGAA 250
67 sGlnThrPheAsnLysValLysTyrTrpLeuGlyAlaArgCysHisIleG 84
251 GCAAACTTTTAAAGTGAAGTGTATTGGGTGCCCGCTCCACATCG 300
84 LuLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrCysSerLysGlu 100
301 AGAAGCTAAAGGAAGTCAATCAGCAGATAAAGATAATATGTCAGTAAGA 350
101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArgSe 117
351 GGCAACTTACTTATTGATGTGGAGCTCTCTCGATCTCAAGGACAACGGAG 400
```



```

167 rAsnValHisPheIleValGlyProProGlyCysGlyLysSerLysTrpA 184
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
551 CAATGTACACGTCTATTGTGGGCCACCTGGTGTGGTAAAGCAAAATGGG 600
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
184 laAlaAsnPheAlaAsnProGluThrThrTyrTrpLysProProLysAsn 200
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
601 CTGCTAATTTTGCAGACCGGAACCACTACTGGAACCACTAGAAAC 650
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
201 LysTrpTrpAspGlyTyrHisGlyLysValValValValLysValLys 217
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
651 AAGTGGTGGGATGTTTACCATGGTGAAGAGTGGTGTGTATGATGACTT 700
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
217 eTyrGlyTrpLeuProTrpAspLeuLeuArgLysCysAspArgTyrP 234
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
701 TTATGGCTGGCTGCGGTGGGATGATCTACTGAGACTGTGTGATCGATATC 750
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
234 roLeuThrValLysThrLysGlyGlyThrValProPheLeuAlaArgSer 250
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
751 CATGTACTGTAGACTAAAGTGAAGTGTACCTTTTGGCCCGCAGT 800
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
251 IleLeuIleThrSerAsnGlnThrProLeuGluTrpTyrSerSerThrAl 267
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
801 ATTCTGATTACCAACATCACACCGGTGGAAATGGTACTCCTCACTGTC 850
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
267 aValProAlaValGluAlaLeuTyrArgArgIleThrSerLeuValPheT 284
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
851 TGTCCAGCTGTAGAAGCTCTCTATCGAGGAGTACTTCTCTGTGTAATTT 900
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
284 rPlysAsnAlaThrLysGlnSerThrGluGluGlyGlyGlnPheValThr 300
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
901 GGAAGAATGCTACAGAAATCCACGAGGAGGAGGCGCCAGTTCGTCAACC 950
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
301 LeuSerProProCysProGluPheProTyrGluIleAsnTyr 314
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
951 CTTTCCCCCATGCCCTGAATTTCCATATGAAATAAATATAC 992
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||

```

seq\_name: gb\_vi:AF166528

```

seq_documentation_block:
LOCUS AF166528 1768 bp mRNA VRL 19-AUG-1999
DEFINITION Porcine circovirus complete genome.
ACCESSION AF166528
NID 95739338
VERSION AF166528.1 GI:57393338
KEYWORDS porcine circovirus.
SOURCE porcine circovirus.
ORGANISM
VIRUSES; ssDNA viruses; Circoviridae; Circovirus.
REFERENCE 1 (bases 1 to 1768)
AUTHORS Yang K.H., Lee Y.F., Chao D.S., Shieh Y.C. and Lai S.S.
TITLE Complete nucleotide sequences of porcine circovirus Tainan strand
OUTBREAK in Taiwan
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1768)
AUTHORS Yang K.H., Lee Y.F., Chao D.S., Shieh Y.C. and Lai S.S.
TITLE Direct Submission
JOURNAL Submitted (08-JUL-1999) Veterinary Medicine, National Chia-Yi
Institute of Technology, 300 Shui Fu Road, Lu Liao Li, Chia-Yi City
600, Taiwan
FEATURES
source
1. .1768
/organism="porcine circovirus"
/virus="porcine circovirus"
/strain="Tainan"
/db_xref="taxon:46221"
/country="Taiwan"
51. .995
/codon_start=1
/product="putative Rep protein"
/protein_id="AAD50432.1"
/db_xref="PID:g5739339"
/db_xref="GI:5739339"
/translation="MPSKNGRSGPQPKRWVTLNPNSEDERKKIRELPISLFDYFI

```

```

VGEGNEGRTPHLOGFANFVKQTFNKVKVFGARCHIEKAKGTDOQNKYCSKEGN
LLIEGAPRSQSDLSSTAVSLTLESGLVTVAGQHPVTFRNFRGLAELLKVSQKM
OKRMDLVNHHVIVGPGGCKSKAANFADPETTYKPRNKWDGYPHGEVVDYDFY
GMLPWDDLRLCDRYPLETVETKGTVPFLARSILLITISNQTPLEWYSSSTAVPAVEALYR
RTISLVFNKATEQSTEGGFVTLSPPCPEPFYELNY"
BASE COUNT 453 a 367 c 492 g 456 t
ORIGIN
alignment_scores:
Quality: 1678.00 Length: 314
Ratio: 5.378 Gaps: 0
Percent Similarity: 99.363 Percent Identity: 97.134
alignment_block:
US-09-209-961-3 x AF166528
Align seg 1/1 to: AF166528 from: 1 to: 1768
1 MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTr 17
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
51 ATGCCCGAGCAAGAAATGGAAGAGCGGCCGCCACCAACCAATAAAGGTG 100
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
17 pValPheThrLeuAsnAsnProSerGluAspLysLysLysLysArgG 34
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
101 GGTGTTCAAGCTCAATAATCTTCCGAAGACGAGCGCAAGAAATAACGGG 150
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
34 luLeuProIleSerLeuPheAspTyrPheIleValGlyGluGluGlyAsn 50
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
151 AGCTCCCAATCTCCCTATTGATTATTATTGTTGGCGAGGAGGTAAAT 200
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
51 GluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPheValLysLy 67
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
201 GAGGAAGCGCAACACCCACCTACAGGGTTCGCTAATTTTGTGAAGAA 250
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
67 sGlnThrPheAsnLysValLysTyrTyrLeuGlyAlaArgCysHisIleG 84
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
251 GCAAACTTTTAAATAAGTGAAGTGTATTTTGGTGGCGCTGCCACATCG 300
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
84 luLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrCysSerLysGlu 100
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
301 AGAAAGCGAAAGGAGCAACATCAGCAGATAAAGAAATACTGCAGTAAAGAA 350
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArgSe 117
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
351 GGCAACTTACTTATCGAGTGTGGAGCTCCTAGATCTCAAGGACACACGAG 400
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValT 134
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
401 TGACCTGTCTACTGCTGTGAGTACTTGTGGAGCGGGAGTCTGTGTGA 450
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
134 hrValAlaGluGlnHisProValThrPheValLysAsnPheArgGlyLeu 150
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
451 CCGTGTGAGAGCAGCACCTCTAAGCTTTGTCAGAAATTTCCCGCGGGCTG 500
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
151 AlaGluLeuLeuLysValSerGlyLysMetGlnLysArgAspTrpLysTh 167
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
501 GCTGAACCTTTTGAAGAGCGGGAATGCAAGACGCTGATTGGAAGAC 550
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
167 rAsnValHisPheIleValGlyProProGlyCysGlyLysSerLysTrpA 184
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
551 CAATGTACACGTCTATTGTGGGCCACCTGGTGTGGTAAAGCAAAATGGG 600
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
184 laAlaAsnPheAlaAsnProGluThrThrTyrTrpLysProProLysAsn 200
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
601 CTGCTAATTTTGCAGACCGGAACCACTACTGGAACCACTAGAAAC 650
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
201 LysTrpTrpAspGlyTyrHisGlyGluLysValValValLysValLys 217
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
651 AAGTGGTGGGATGTTTACCATGGTGAAGAGTGGTGTGTATGATGACTT 700
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
217 eTyrGlyTrpLeuProTrpAspLeuLeuArgLysCysAspArgTyrP 234
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||

```

```
701 TTATGTTGGCTGCCGTGGATGATCTACTGAGACTGTGTGATGCATATC 750
234 rleuThrVallyThrlyGlyThrValProPheLeuAlaArgSer 250
|||||:::|||||
751 CATTCAGCTAGAGCTAAGGTGAAGTGTACCTTTTGGCCCGCAGT 800
251 ileLeuileThrSerAsnGlnThrProLeuGluIuTrpTyrSerSerThrAl 267
|||||:::|||||
801 ATTCTGATTACCAGCAATCAGACCCGTTGGAATGGTACTCTCAACTGC 850
267 aValProAlaValGluAlaLeuTyrArgArgileThrSerLeuValPheT 284
|||||:::|||||
851 TGTCACAGCTAGAGCTCTCTATCGGAGGATTACTTCTCTGGTATTT 900
284 rPLySAsnAlaThrlySGlnSerThrGluGluGlyGlyGlnPheValThr 300
|||||:::|||||
901 GGAAGAATCTACAGCAATCCAGGAGGAGGGCGAGTCGTCACC 950
301 LeuSerProCysProGluPheProTyrGluIleAsnTyr 314
|||||:::|||||
951 CTTTCCCCCCTGCTGCTGATTTCCATATGAATAAATTAC 992
seq_name: gb_vi:AF109399
seq_documentation_block: 1768 bp DNA circular VRL 06-JAN-1999
LOCUS AF109399 Porcine circovirus type 2-E, complete genome.
DEFINITION Porcine circovirus type 2-E, complete genome.
ACCESSION AF109399
NID 94106896
VERSION AF109399.1 GI:4106896
KEYWORDS
SOURCE porcine circovirus type 2-E.
ORGANISM porcine circovirus type 2-E.
VIRUSES; ssDNA viruses; Circoviridae; Circovirus; #59#line
REFERENCE 1 (bases 1 to 1768)
AUTHORS Hamel,A.L. and Nayar,G.P.S.
TITLE Nucleotide sequence of four different isolates of circovirus
detected in pigs with various clinical syndromes
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1768)
AUTHORS Hamel,A.L. and Nayar,G.P.S.
TITLE Direct Submission
JOURNAL Submitted (27-NOV-1998) Virology Laboratory, Veterinary Diagnostic
Laboratory, Manitoba Agriculture, 545 University Crescent,
Winnipeg, Manitoba R3T 5S6, Canada
FEATURES
source
1..1768
/organism="porcine circovirus type 2-E"
/db_xref="taxon:85544"
/notes="similar to Porcine circovirus sequence presented in
GenBank Accession Number AF027217; type-E designation is
based upon restriction endonuclease digestion pattern;
sequence obtained from several overlapping PCRs using DNA
extracted from lung, mesenteric lymph node and tonsil of
pig"
repeat_region 13..35
/rpt_type=tandem
/rpt_unit=13..18
51..995
/notes="ORF-1."
/codon_start=1
/product="putative Rep and coat protein"
/protein_id="AAD03071.1"
/db_xref="PID:g4106897"
/db_xref="GI:4106897"
/translation="MPSKNGRSGPQPKRWFTLNNPSEDEKKIRELPISLFDYFI
VGEENEGRTPLHGFANFKQTKNKKVILGACGHIKAGTDOONKEYCSKEGN
LLIECAPRSOGORSLSLSTAVSTLLESGSLTVVAEOHPVTFVRNFRGLAELLKVSQGM
OKRDKNTNHIIVGPPCGCKSKWANFADPTTYWPKPNKWDGYPHGEVVDYDFY
GILPWDDLRLCRLDRPLTVETKGTVPFLARSILITSNQPLEWYSPSTAVPAVEALYR
RITSLVFWKNATEQSTEEGGQFVTLSPPCPEFFYEINY"
polyA_signal
CDS
complement(357..671)
/codon_start=1
/product="ORF-3"
/protein_id="AAD03073.1"
/db_xref="PID:g4106899"
/db_xref="GI:4106899"
/translation="MVTIPPLVSRWFPVCGFRVCKISSPFAFTTTPRWPHNDVY IGLPI
TLHFPAHFQKFSQPAEISDKRYRVLLNGHOTPALQOQTHSSRQVTPLSRSSTL
NK"
complement(386..565)
/codon_start=1
/product="ORF-4"
/protein_id="AAD03074.1"
/db_xref="PID:g4106900"
/db_xref="GI:4106900"
/translation="MTCITLVFQSRFCIFPLTFKSSASPRKFLINVTGCCSATVTRLPL
SNKVLTAVDRLSLRCP"
553..732
/codon_start=1
/product="ORF-12"
/protein_id="AAD03081.1"
/db_xref="PID:g4106907"
/db_xref="GI:4106907"
/translation="MYTSLWGLGVVVKANGLLILQTRKPHGTGNHLETSGGMVTVMVKKW
LLMTFMAGCRGMIY"
complement(688..753)
/codon_start=1
/product="ORF-8"
/protein_id="AAD03078.1"
/db_xref="PID:g4106904"
/db_xref="GI:4106904"
/translation="MDIDHTVSDVHDPTAASHKSHQ"
983..988
complement(989..1033)
/codon_start=1
/product="ORF-11"
/protein_id="AAD03080.1"
/db_xref="PID:g4106906"
/db_xref="GI:4106906"
/translation="MNNKNHYEVKKTKQ"
1016..1363
/notes="similar to Porcine circovirus ORF-5 encoded by the
sequence presented in GenBank Accession Number AF027217."
/codon_start=1
/product="ORF-5"
/protein_id="AAD03075.1"
/db_xref="PID:g4106901"
/db_xref="GI:4106901"
/translation="MVFILHGFKNVFKIFSELYIHGYTDIVLVVFTVPERSAEA
YVHISLGLESHPOLIPFVILWLEINSIGIKRFGCEVPVGVGEGGLDCMAGGVVYVG
IGLGGLYYKVIY"
complement(1022..1027)
complement(1034..1735)
/codon_start=1
/product="ORF-2"
/protein_id="AAD03072.1"
/db_xref="PID:g4106898"
/db_xref="GI:4106898"
/translation="MTYPRRRRRRRHPRSHLQGLIRRRPWLHPHRYWPKRKNIGI
FNARLSRTFGITVATPSTPSWAVMLRNLDDFPVPGGGNKISIFEFYIRIKRYK
VEFWPCSPITQDGRGVSSAILDDNFVIKATQIDYDYNYSRSHYIPQPFYSHSRY
FTPKPVLDSITIDYFQPNKRNQLWMRLQTSRNVHDVHGLGTAFENSKYQDQYNIRVTMY
VQFREFNLKDPPLKP"
1524..1631
/codon_start=1
/product="ORF-10"
/protein_id="AAD03079.1"
/db_xref="PID:g4106905"
/db_xref="GI:4106905"
/translation="MSTAQEGLTVVLTVPYKVRERALKMPFLFOR"
complement(1528..1611)
/codon_start=1
CDS
```

```
/product="ORF-6"
/protein_id="AAD03076.1"
/db_xref="PID:g4106902"
/db_xref="GI:4106902"
/translation="MASSTPASPAPSDILSRLPQARPGR"
1682..1741
CDS
/codon_start=1
/product="ORF-7"
/protein_id="AAD03077.1"
/db_xref="PID:g4106903"
/db_xref="GI:4106903"
/translation="MAAGAVSSAETPPWIRHS"
join(1750..1768,1..13)
join(1762..1768,1..2)
/stem_loop
rep_origin
/note="putative; similar to the nonanucleotide motif of
Porcine circovirus presented in GenBank Accession Numbers
AF027217 and U9186"
BASE COUNT 448 a 362 c 495 g 463 t
ORIGIN

alignment_scores:
  Quality: 1677.00      Length: 314
  Ratio: 5.392          Gaps: 0
  Percent Similarity: 99.045  Percent Identity: 97.134

alignment_block:
US-09-209-961-3 x AF109399 ..
Align seg 1/1 to: AF109399 from: 1 to: 1768
1 MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTr 17
51 ATGCCAGCAGAGAGATGAAGAGCGGCGCCCAACCAACCAATAAAGGTG 100
17 pValPheThrLeuAsnAsnProSerGluAspGluArgLysLysileArg 34
101 GGTCTTCACGCTGAATAATCTCCGAAGACGAGCGCAAGAAATACGGG 150
34 LuLeuProLieserLeuPheAspTrpPheIleValGlyGluGlyAsn 50
151 AGCTTCCAATCTCCCTGTTGATTATTATTGTTGGCAGGAGGTAAT 200
51 GluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPheValLys 67
201 GAGGAAGACGACCAACCTCCACCTCAGGGTTCCTAATTTGTGAAAA 250
67 sGlnThrPheAsnLysValLysTrpTrpLeuGlyAlaArgCysHisIle 84
251 GCAAACTTTTAAATAAGTGAAGTGTATCTTTGTTGGTCCGCACATCG 300
84 LuLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrcysSerLysGlu 100
301 AGAAGGCAAGGAAGTATGATCAGCAGATAAAGAAATATTGCAGTAAGA 350
101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArg 117
351 GGCAACTTACTTATGAGTGGAGGACCTCGATCTCAAGACACCGGAG 400
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuVal 134
401 TGACCTGCTACTCTGTGAGTACTCTGTTGGAGCGGAGTCTGGTGA 450
134 hrValAlaGluGlnHisProValThrPheValLysAsnPheArgGlyLeu 150
451 CCGTGTGACAGACGACCCCTTAACGTTGTCAGAAATTTCCCGCGGGTG 500
151 AlaGluLeuLeuLysValSerGlyLysMetGlnLysArgAspTrpLys 167
501 GCTGAACCTTTTGAAGTGACGGGAAAATGCAAGCGGTGATTCGAAGAC 550
167 rAsnValHisPheIleValGlyProGlyCysGlyLysSerLysTrp 184
|||||
```

```
551 CAATGTACACGTCATTTGTGGGCCACCTGGGTGGTAAAGCAATGGG 600
184 laAlaAsnPheAlaAsnProGluThrThrTyrrTpLysProProLysAsn 200
601 CTGCTAATTTTGCAGACCCGGAACCACTACTGGAACACCACTAGAAC 650
201 LysTrpTrpAspGlyTyrrHisGlyGluLysValValIleAspAspPh 217
651 AAGTGGTGGGATGTTACCATGGTGAAGAGTGGTTGTTATGATGACT 700
217 eTyrrGlyTrpLeuProTrpAspAspLeuLeuArgLeuCysAspArgTy 234
701 TTATGGTGGTGGTGGGATGATCTACTGAGACTGTGTGATCGATATC 750
234 rLeuThrValLysThrLysGlyThrValProPheLeuAlaArgSer 250
751 CATGTACTGTAGAGACTTAAGGTGAAGTGTACCTTTTGGCCCGCAGT 800
251 IleLeuIleThrSerAsnGlnThrProLeuGluTrpTyrrSerThrAl 267
801 ATTCTGATTACCAGCAATCAGACCCCGTTGGAATGGTACCCCTCACT 850
267 aValProAlaValGluAlaLeuTyrrArgIleThrSerLeuValPheT 284
851 TGTCCAGCTGTAGAAGCTCTCTATCGGAGGATTACTTCTTGGTATT 900
284 rPlysAsnAlaThrLysGlnSerThrGluGlyGlyGlnPheValThr 300
901 GGAAGATGCTACAGAACATCCAGGAGGAGGGGCCAGTTCGTCCACC 950
301 LeuSerProProCysProGluPheProTyrrGluIleAsnTrp 314
951 CTTTCCCCCATGCCCTGAATTTCCATATGAATAAATTAC 992

seq_name: gb_vi:AF055393
seq_documentation_block:
LOCUS AF055393 1767 bp DNA circular VRL 13-SEP-1998
DEFINITION Porcine circovirus Type II from France, complete genome.
ACCESSION AF055393
NID 93598820
VERSION AF055393.1 GI:3598820
KEYWORDS porcine circovirus.
SOURCE Porcine circovirus.
ORGANISM Viruses; SSDNA viruses; Circoviridae; Circovirus.
REFERENCE 1 (bases 1 to 1767)
AUTHORS Meehan,B.M., McNeilly,F., Todd,D., Kennedy,S., Jewhurst,V.A.,
Ellis,J.A., Hassard,L.E., Clark,E.G., Haines,D.M. and Allan,G.M.
TITLE Characterization of novel circovirus DNAs associated with wasting
syndromes in pigs
JOURNAL J. Gen. Virol. 79 (Pt 9), 2171-2179 (1998)
MEDLINE 98418498
REFERENCE 2 (bases 1 to 1767)
AUTHORS Meehan,B.M.
TITLE Direct Submission
JOURNAL Submitted (26-MAR-1998) Brian M Meehan, Veterinary Science, The
Queen's University of, Belfast, Stoney Road, Belfast, Northern
Ireland, BT4 3SD, U.K.
FEATURES
source
1..1767
/organism="porcine circovirus"
/specific_host="Sus scrofa"
/strain="Porcine circovirus Type II"
/db_xref="taxon:46221"
/clone="Imp.1011 48121 ECORI No.17"
/country="France"
/complement(101..190)
/codon_start=1
/product="ORF6"
/protein_id="AAC35329.1"
/db_xref="PID:g3598830"
/db_xref="GI:3598830"
CDS
```



```

/translation="MASSTPASPAPSDILSSPEQSERPPGRWT"
103..210
/codon_start=1
/product="ORF10"
/protein_id="AAC35327.1"
/db_xref="PID:g3598828"
/db_xref="GI:3598828"
/translation="MSTAQEGVLTVVRLTVYPKVRRRLVKMPFFLLQR"
261..320
/codon_start=1
/product="ORF7"
/protein_id="AAC35328.1"
/db_xref="PID:g3598829"
/db_xref="GI:3598829"
/translation="MAAGAVSSSPVTPPWIRHI"
complement(311..439)
/notes="Predicted 4.6 kDa protein"
/codon_start=1
/product="ORF9"
/protein_id="AAC35325.1"
/db_xref="PID:g3598826"
/db_xref="GI:3598826"
/translation="MGLGSASSILLARHVAEVLPRCCRCRSALVILTAHFRRFQI"
398..1342
/codon_start=1
/product="putative Rep protein (ORF1)"
/protein_id="AAC35320.1"
/db_xref="PID:g3598821"
/db_xref="GI:3598821"
/translation="MPSKNGRSGPQPKRWVFTLNNPSEDERKKIRDLPLISLDFYFI
VDEGEDEGRTHLQGFANFVKOTENKVKYLGARCHEKAKGTQONKEYCKSGKN
LMECCGAPSGQSDLSYAVSLTLESGSLVTVAEOHPVTFVRFNGLAELLKVKSKM
QKRDWTNHYIVGPGCCGSKSWAANFADPTIYKPPNKNWWDGTHGEVYVVDIFY
GWLPMDDLRLCDGLYVEITKGTVPFLARSILITSNQTPLEWISSTAVPAVEALYR
RITSLVFNKNAEQSTEEGQFVTLSPPCPEPFYEINY"
674..679
complement(704..1018)
/notes="predicted 11.9 kDa protein"
/codon_start=1
/product="ORF3"
/protein_id="AAC35322.1"
/db_xref="PID:g3598823"
/db_xref="GI:3598823"
/translation="MYTIPLYSRWFPGVGRVCKISSPFAFTTPRWPHNDVYISLPI
TLHFFPAHQKFSQPAEISDKRYRVLNCGHQTALQQGTHSSRQVTPILSRSRSTL
HQ"
complement(733..912)
/notes="predicted 6.5 kDa protein"
/codon_start=1
/product="ORF4"
/protein_id="AAC35323.1"
/db_xref="PID:g3598824"
/db_xref="GI:3598824"
/translation="MTCTLVFQSRFCIFPLTFKSSASPRKFLTNVTCGSATVTRPLP
SNKVLTAVDRLRCP"
complement(1035..1100)
/notes="predicted 2.3 kDa protein"
/codon_start=1
/product="ORF8"
/protein_id="AAC35324.1"
/db_xref="PID:g3598825"
/db_xref="GI:3598825"
/translation="MDIDHTSVVDHPRAASHKSHQ"
1330..1335
complement(1336..1380)
/notes="predicted 1.8 kDa protein"
/codon_start=1
/product="ORF11"
/protein_id="AAC35326.1"
/db_xref="PID:g3598827"
/db_xref="GI:3598827"
/translation="MNNKNHYEVIKTKQ"
complement(1369..1374)

```

CDS

complement(join(1380..1767,1..314))  
/notes="predicted 27.8 kDa protein"  
/codon\_start=1  
/product="ORF2"  
/protein\_id="AAC35321.1"  
/db\_xref="PID:g3598822"  
/db\_xref="GI:3598822"  
/translation="MTYPRRYYRRRRHPRSHLQILRRRPWLVPYHRRYWRKNGI  
ENTRLSRTFGYTVKRTTVPSPVAVDMMREINDFLPPGGSNPRSPYFEYIRKVK  
VEFPCSPITQDGRGVGSSAVILDDNFYTKATALTDPYVNYSSRHITQPFYSYSKY  
FTPKPVLDSIDIFYQPNKRNQLWLRLQTAGNVHDVGLGAFENSIYDQETNIRVMT  
VOFRENLKDPPLNP"

BASE COUNT 447 a 360 c 502 g 458 t  
ORIGIN

alignment\_scores:  
Quality: 1676.00 Length: 314  
Ratio: 5.372 Gaps: 0  
Percent Similarity: 99.363 Percent Identity: 96.815

alignment\_block:  
US-09-209-961-3 x AF055393 ..

Align seg 1/1 to: AF055393 from: 1 to: 1767

1 MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTr 17  
|||||  
398 ATGCCGACGAAGAAGAAATGGAGAAGCGGACCCCAACCCCATAAAGGTG 447  
|||||

17 pValPheThrLeuAsnAsnProSerGluAspGluArgLysLysIleArgG 34  
|||||  
448 GGTGTCTACTCTGAATATCTCTCGAGACGAGCGGCAAGAAATACGGG 497  
|||||

34 LuLeuProIleSerLeuPheAspTyrPheIleValGlyGluGlyAsn 50  
|||||  
498 ATCTTCCAATATCCCTATTATTGATTATTTATTGTTGGCGAGGAGGTAA 547  
|||||

51 GluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPheValLysL 67  
|||||  
548 GAGGAAGACGACCAACACCTCACCTCCAGGGGTTTCGCTAATTTTGTGAAGAA 597  
|||||

67 sGlnThrPheAsnLysValLysTyrTyrLeuGlyAlaArgCysHisIleG 84  
|||||  
598 GCAGACTTTTATAAAGTGAAGTGTATTGTTGGTCCCTCCCATCATCG 647  
|||||

84 LuLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrCysSerLysGlu 100  
|||||  
648 AGAAAGCGAAAGGAACAGATCAGCAGATAAAGATACTGCAAGTAAAGAA 697  
|||||

101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlnArgSe 117  
|||||  
698 GGCAACTTACTGATGAGTGTGGAGCTCTCTAGATCTCAGGGACAACGGAG 747  
|||||

117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValT 134  
|||||  
748 TGACCTGTCTACTGTGTGAGTACCTTGTGGAGAGCGGGAGCTCTGGTGA 797  
|||||

134 hrValAlaGluGlnHisProValThrPheValLysAsnPheArgGlyLeu 150  
|||||  
798 CCGTTGCAGACGACGACCTCTAAGCTTTGTGAGAAATTCCTCCGGGCTG 847  
|||||

151 AlaGluLeuLeuLysValSerGlyLysMetGlnLysArgAspTrpLysTh 167  
|||||  
848 GCTGAACATTTTAAAGTGAAGGAAATGCAGAAAGCGGTGATTGGAAGAC 897  
|||||

167 rAsnValHisPheIleValGlyProProGlyCysGlyLysSerLysTrpA 184  
|||||  
898 TAATGTACACGTCATTTGTGGGCCACCTGGGTGTGTGTAAGAAACAATGGG 947  
|||||

184 laAlaAsnPheAlaAsnProGluThrThrTyrTrpLysProProLysAsn 200  
|||||  
948 CTGCTAAATTTGCAGACCCGCGGAAACACATACTGGAACACCACCTAGAAAC 997  
|||||

polyA\_signal  
CDS

polyA\_signal  
CDS

polyA\_site



/db\_xref="PID:g3598833"  
/db\_xref="GI:3598833"  
/translation="MTYPRRRRRRRPRSHLQIILRRPWLVPRRHYRWRKNGI  
FNRLSRTFGYTKRTTKTPSWADVMRENINDLPPGGSNPRSPFFEYIRKVK  
VEWPSPITQGRGVSSAVIDDNFTKATALTYPDYNYSRRHITQPFVSHRY  
FTKPVLDSTYDFPNKRNQWLRLQTTGNVDHVLGTAFENSIDQENIRVTMY  
VQPREFNKDPPLNP"  
BASE COUNT 448 a 359 c 500 g 460 t  
ORIGIN

alignment\_scores:  
Quality: 1676.00 Length: 314  
Ratio: 5.372 Gaps: 0  
Percent Similarity: 99.363 Percent Identity: 96.815

alignment\_block:  
US-09-209-961-3 x AF055394 ..

Align seg 1/1 to: AF055394 from: 1 to: 1767

1 MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTr 17  
398 ATGCCAGCAAGAAGAAATGGAAGAGCGGACCCCAACCCCAATAAAGGTG 447  
17 pValPheThrLeuAsnAsnProSerGluAspGluArgLysLysLleArgG 34  
448 GGTGTCTACTCTGAATAATCCCTCCGAGACGAGCGCAAGAAATACGGG 497  
34 LuLeuProIleSerLeuPheAspTyrPheIleValGlyGluGluGlyAsn 50  
498 ATCTCCATATCCCTATTGTGATTATTTATTGTTGGCGAGGAGGTAT 547  
51 GluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPheValLysly 67  
548 GAGGAAGGAGCAACACCTCACCTCCAGGGGTTCTGCTAATTTGTGAAGAA 597  
67 sGlnThrPheAsnLysValLysTrpTyrLeuGlyAlaArgCysHisIleG 84  
598 GCAGACTTTTATAAGTGAAGTGTATTGTTGGTGGCCCTGCCACATCG 647  
84 LuLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrCysSerLysGlu 100  
648 AGAAGCGAAGAACACAGATACAGCAGATAAAGATATCTGCAGTAAAGAA 697  
101 GlyAsnLeuLeuIleGlyCysGlyAlaProArgSerGlnGlyLysArgse 117  
698 GCCAACTTACTGAGTGTGGAGTCTCTAGATCTCAGGACACAGGAG 747  
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyLleLeuValT 134  
748 TGACCTGTCTACTGTGTGAGTACCTTGTGGAGACGGGAGTCTGGTGA 797  
134 hrValAlaGluGlnHisProValThrPheValLysAsnPheArgGlyLeu 150  
798 CCGTGTGACAGCAGACCCCTGTAACGTTTGTGAGAAATTTCCCGGGCGTG 847  
151 AlaGluLeuLeuLysValSerGlyLysMetGlnLysArgAspTrpLysTh 167  
848 GCTGAACCTTTGAAGTGAAGCGGAAATGCAGACCGTGTGTTGGAAGAC 897  
167 rAsnValHisPheIleValGlyProProGlyCysGlyLysSerLysTrpA 184  
898 TAATGTACACGTCATTGTGGGCCACCTGGTGGTGTGTTGTTGAAAGCAATGG 947  
184 laAlaAsnPheAlaAsnProGluThrThrTyrTrpLysProProLysAsn 200  
948 CTGCTAATTTTCAGACCGGAAACACACATCTGGAACACCACTAGAAAC 997  
201 LysTTrpTrpAspGlyTyrHisGlyGluLysValValAlaIleAspAspPh 217  
998 AAGTGTGGGATGGTGTACCATGGTGAAGAAGGTGTGTTGTTATGATGACTT 1047

217 eTyrGlyTrpLeuProTrpAspLeuLeuArgLeuLeuCysAspArgTyrP 234  
1048 TTATGGCTGGCTGCCCTGGATGATCTACTGAGACTGTGTGATCGATATC 1097  
234 rLeuThrValLysThrLysGlyGlyThrValProPheLeuAlaArgSer 250  
1098 CATTGACTAGAGACTAAAGGTGGAACGTGTACTTTTGTGGCCCGCAGT 1147  
251 IleLeuIleThrSerAsnGlnThrProLeuGluTrpTyrSerSerThrAl 267  
1148 ATTCTGATTACCAAGCAATCAGACCCGCTTGGAAATGGTACTCCTCAACTGC 1197  
267 aValProAlaValGluAlaLeuTyrArgArgIleThrSerLeuValPheT 284  
1198 TGTCCTCAGCTGTAGAAGCTCTTTATCGGAGGATTACTTCTCTGGTATTT 1247  
284 rPlysAsnAlaThrLysGlnSerThrGluGluGlyGlyGlnPheValThr 300  
1248 GGAAGAATGCTACAGAAACATCCACGAGGAGGGGCCAGTTTCGTCAAC 1297  
301 LeuSerProProCysProGluPheProTyrGluIleAsnTyr 314  
1298 CTTTCCCCCATGCTGCTGCTGAATTTCCATATGAAATAAATTAC 1339

seq\_name: gb\_v1:AF027217

seq\_documentation\_block:

LOCUS AF027217 1768 bp DNA circular VRL 14-MAY-1998  
DEFINITION Porcine circovirus strain pmws PCV, complete genome.

ACCESSION AF027217  
NID 92689645  
VERSION AF027217.1 GI:2689645

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

AUTHORS

TITLE

JOURNAL

Manitoba R3T 5S6, Canada

Location/Qualifiers

1. 1768

/organism="porcine circovirus"

/strain="pmws PCV"

/db\_xref="taxon:46221"

/note="both strands of seven overlapping PCR fragments

were sequenced; virus isolated from lung, lymph node,

spleen and tonsil tissue from pigs affected by post

weaning multisystemic wasting syndrome"

51. 995

/note="ORF1: similar to Rep protein encoded by

non-pathogenic PCV, GenBank Accession Number U49186;

predicted 35.8 kDa protein"

/codon\_start=1

/product="putative Rep protein"

/protein\_id="AAC59462.1"

/db\_xref="PID:g2689646"

/db\_xref="GI:2689646"

/translation="MPSKKNRSGPQPKHRVFTLNNPSEDERKKIRELPISLDFYFI

VGEENEGRTPLHGFANFYKQTFNKVKWYLGARCYIEKAKGTDQNKCYCSKEGN

LLIECGAPRSQQRSDISTAVSTLLESGSLVTVAEQHPVTFRNFRGLAEILKVGSKM

OKRDWNTNHHVIGVPGCGSKWAANFADPTTYWKPRNKWDCYHGEEVVVDDFY

GWLPWDDLRLICDRSYPILTVEKTKGTVPFLARSLITSNQTPLEWYSSATVPAVEALYR

RTSLVFWKNATEGTEGGGFVTLSPPCPEPPEINT"

117. 125

misc\_feature

```
polyA_signal
CDS
/note="glycosylation site"
327..332
complement(357..671)
/note="ORF3; predicted 11.9 kDa protein"
/codon_start=1
/protein_id="AAC59464.1"
/db_xref="PID:g2689648"
/db_xref="GI:2689648"
/translation="MVTIPLYSRWFPVCGFRVCKISSPPAETTPRWPHNDVYIGLPI
TLLHPAPHQKFSQPAESDKRYRVLLCNGHQTALQOQTHSSRQVTPLSLRSSTF
NK"
complement(386..565)
/note="ORF4; predicted 6.5 kDa protein"
/codon_start=1
/protein_id="AAC59465.1"
/db_xref="PID:g2689649"
/db_xref="GI:2689649"
/translation="MTCTLVFQSRFCIFPLTFKSSASPRKFLTNVTGCCSATVTRLPL
SNKVLTAVDRLSRCP"
complement(470..478)
/note="glycosylation site"
complement(688..753)
/note="ORF8; predicted 2.3 kDa protein"
/codon_start=1
/protein_id="AAC59469.1"
/db_xref="PID:g2689653"
/db_xref="GI:2689653"
/translation="MDIDHTVSDHPTAASHKSHQ"
816..824
/note="glycosylation site"
906..914
/note="glycosylation site"
983..988
complement(989..1033)
/note="ORF11; predicted 1.8 kDa protein"
/codon_start=1
/protein_id="AAC59472.1"
/db_xref="PID:g2689656"
/db_xref="GI:2689656"
/translation="MKNKHVEYIKTKQ"
1016..1177
/note="ORF5; predicted 6.2 kDa protein"
/codon_start=1
/protein_id="AAC59466.1"
/db_xref="PID:g2689650"
/db_xref="GI:2689650"
/translation="MVFIRPHLGKMGVFKIFSELYIHGYTDIVLVVYTFERSAEA
YVYHISRLGI"
complement(1022..1027)
complement(1034..1735)
/note="ORF2; predicted 27.8 kDa protein"
/codon_start=1
/protein_id="AAC59463.1"
/db_xref="PID:g2689647"
/db_xref="GI:2689647"
/translation="MTYPRRYRRRRHPRSHLQILRRPWLHPHRYRNRKNGI
FNILSRFGYTVKATVTPSWAVDMRFNIDDFVPPGGGNKISPEYIRIRKVK
VFWRPCSPQTQDGVGSTAVILDDNFVKATALYDPYVNYSSRHTIPQPFYSHSRY
FTPKVLDSTIDYFQPNKRTQLWLRLQTSRNVHDVHGLTAFENSIIYDQYIRVTMY
VQFREFNLKDPPLKP"
complement(1301..1309)
/note="glycosylation site"
complement(1522..1611)
/note="ORF6; predicted 3.1 kDa protein"
/codon_start=1
/protein_id="AAC59467.1"
/db_xref="PID:g2689651"
/db_xref="GI:2689651"
/translation="MASSTPASPAPSDILSRLPQSERPPGRWT"
1524..1631
/note="ORF10; predicted 4.1 kDa protein"
/codon_start=1
/protein_id="AAC59471.1"
```

```

/db_xref="PID:g2689655"
/db_xref="GI:2689655"
/translation="MSTAQEGVLTVALTVYKVRERRVLMKPFLLQR"
1682..1741
/note="ORF7; predicted 1.9 kDa protein"
/codon_start=1
/protein_id="AAC59468.1"
/db_xref="PID:g2689652"
/db_xref="GI:2689652"
/translation="MAAGAVSSSAVTPPWIRHS"
complement(1732..1768.1..92))
/note="ORF9; predicted 4.6 kDa protein"
/codon_start=1
/protein_id="AAC59470.1"
/db_xref="PID:g2689654"
/db_xref="GI:2689654"
/translation="MWLGSSAILAGHVAAEVLPRCCRCRSALVILTAHFFRFQL"
Join(1746..1768.1..113)
/function="putative replication site"
Join(1762..1768.1..2)
/note="AAGTATTAC" is similar to the nonanucleotide motif
in the non-pathogenic PCV, GenBank Accession Number
U49186"
BASE COUNT 452 a 360 c 495 g 461 t
ORIGIN

alignment_scores:
  Quality: 1676.00      Length: 314
  Ratio: 5.372          Gaps: 0
  Percent Similarity: 99.363  Percent Identity: 97.134

alignment_block:
US-09-209-961-3 x AF027217

Align seg 1/1 to: AF027217 from: 1 to: 1768

1 MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTr 17
|||||
51 ATGCCCAAGCAGAAAGATGGAAGAACGGACGCCCAACACATAAAGGTG 100
|||||
17 pValPheThrLeuAsnAsnProSerGluAspGluArgLysLysIleArg 34
|||||
101 GGTTTCACGCTGAATATCCTCCGAAGACGCGCAAGAATAACGG 150
|||||
34 luLeuProIleSerLeuPheAspTyrPheIleValGlyGluGlyAsn 50
|||||
151 AGCTCCCAATCTCCCTATTGATTATTTATTGTTGGCAGAGGGTAAT 200
|||||
51 GluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPheValLys 67
|||||
201 GAGGAAGGACGACACCTCACCCTCCAGGGGTCGCTAATTTGTGAAGA 250
|||||
67 sGlnThrPheAsnLysValLysTyrTyrLeuGlyAlaArgCysHisIle 84
|||||
251 GCAACATTTAATAAAGTGAAGTGGTATTGGGTGCCCGCTGCTACATCG 300
|||||
84 luLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrCysSerLysGlu 100
|||||
301 AGAAAGCAAAGGAACGTATCAGCAGAAATAAAGATATTTCAGTAAAGAA 350
|||||
101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArgSe 117
|||||
351 GGCAACTTACTTATTGAATGTGGAGCTCCTCGATCTCAAGGACACCGAG 400
|||||
117 rAspLeuSerThrAlaValSerThrLeuGluSerGlyIleLeuValT 134
|||||
401 TGACCTTGCTACTGCTGTGAGTACTTGTGGAGCGGGAGTCTGGTGA 450
|||||
134 hrValAlaGluGlnHisProValThrPheValLysAsnPheArgGlyLeu 150
|||||
451 CCGTTGCAGACGACCCCTGTAACGTTTGTCAAAATTTCCGCGGGCTG 500
|||||
```

```
151 AlaGluLeuLysValSerGlyLysMetGlnLysArgAspTrpLysTh 167
|||||
501 GCTGAACCTTTCAAGAGTACGGGAAATGCAGAACGCTGATTGGAAGAC 550
|||||
167 rAsnValHisPheIleValGlyProGlyCysGlyLysSerLysTrpA 184
|||||
551 CAATGTACACGTCATTGTGGGCCACCTGGGTGGTAAAGCAATGGG 600
|||||
184 IaAlaAsnPheAlaAsnProGluThrThrTyrTrpLysProLysAsn 200
|||||
601 CTGCTAATTTTCACACCGGAAACACATCTGGAACACCACTAGAAAC 650
|||||
201 LysTrpTrpAspGlyThrHisGlyGluLysValValIleAspAspPh 217
|||||
651 AAGTGGTGGATGGTTACCATGTTGAAGAAGTGGTTGTTATTGATGACT 700
|||||
217 eYrGlyTrpLeuProTrpAspLeuLeuArgLeuCysAspArgTyrP 234
|||||
701 TTATGGCTGGCTGCCGTGGATGATCTACTGAGACTGTGTGATCGATATC 750
|||||
234 rGluThrValLysThrLysGlyThrValProPheLeuAlaArgSer 250
|||||
751 CAATGACTAGAGACTAAAGTGGGAAGTGTACCTTTTGGCCCGCAGT 800
|||||
251 IleLeuIleThrSerAsnGlnThrProLeuGluTrpTyrSerSerThrAl 267
|||||
801 ATTCGATTACAGCAATCAGACCCGGTGGAGTGGTACTCTCACTGC 850
|||||
267 aValProAlaValGluAlaLeuTyrArgArgIleThrSerLeuValPheT 284
|||||
851 TGTCCAGCTGTAGAAGCTCTCTATCGGAGGATTACTTCTTGGTATTT 900
|||||
284 rPlyAsnAlaThrLysGlnSerThrGluGlyGlyGlnPheValThr 300
|||||
901 GGAAGAATGCTACAGAACAATCCAGGAGGAAGGGGCCAGTTCGTCAAC 950
|||||
301 LeuSerProProCysProGluPheProTyrGluIleAsnTyr 314
|||||
951 CTTTCCCCCGGCTGCTGAATTTCCATATGAATAAATTAC 992
```

seq\_name: gb\_v1:AF117753

```
seq_documentation_block: 1768 bp DNA circular VRL 04-FEB-1999
LOCUS AF117753
DEFINITION Porcine circovirus type 2-D, complete genome.
ACCESSION AF117753
NID 94219093
VERSION AF117753.1 GI:4219093
KEYWORDS
SOURCE
  ORGANISM
    porcine circovirus type 2-D.
    porcine circovirus type 2-D
    Viruses; ssDNA viruses; Circoviridae; Circovirus; #;#;ine
    Circovirus type 2.
  1 (bases 1 to 1768)
  Hamel,A.L. and Nayar,G.P.S.
  Genetic characterization of four novel type-2 Porcine circoviruses
  JOURNAL Unpublished
  2 (bases 1 to 1768)
  Hamel,A.L. and Nayar,G.P.S.
  Direct Submission
  JOURNAL Submitted (05-JAN-1999) Virology Laboratory, Manitoba Agriculture,
  545 University Crescent, Winnipeg, Manitoba R3T 5S6, Canada
  Location/Qualifiers
  1..1768
    /organism="porcine circovirus type 2-D"
    /virus
    /db_xref="taxon:86385"
    /note="type 2-D designation is based upon restriction
    endonuclease digestion pattern
    several overlapping PCR fragments were sequenced; virus
    isolated from lung, mesenteric lymph node and tonsil
    tissue; similar to Porcine circovirus sequence presented
    in GenBank Accession Number AF027217"
```

```
repeat_region
13..35
/rpt_type=tandem
/rpt_unit=13..18
51..995
/note="putative Rep and coat protein"
/codon_start=1
/product="ORF-1"
/db_xref="PID:94219094"
/db_xref="GI:4219094"
/translation="MPSKNGRSGPQPHKRWFTLNNPSEDERKKIRELPISLFDYFI
VGEENEGRTPHLQGFANFKVQTFNKVYLGARCHIEKAGTDQONKEYCSKEGN
LLMECCAPRSQGSDDLSTAVSTLLESGSLVTAQHPVTYVFRNFRGLAEKLVSKGM
OKRWKTNVHVIVGPPGGCKSKAANFADPTTYKPKRNKWDGYHGVEVVIDDFY
GWLWDLLRLCDRYPLAVETKGTVPPLARSILITSNOTPLEWISSAAPPAVEALYR
RITSLVFWKNATEQSTEGGGFVLSPPCPPEPPEYIN"
327..332
complement(357..671)
/codon_start=1
/product="ORF-3"
/protein_id="AAD12309.1"
/db_xref="PID:94219095"
/db_xref="GI:4219095"
/translation="MVTIPPLVSRWFPVCGFRVCKISSPFAFATPRWPHNDVYIRLPI
TLLHFPAHFQKSPQAISDKRYRVLLCNGHOTLPALKQTHSGROVTPLSLRSSSTF
HQ"
complement(386..565)
/codon_start=1
/product="ORF-4"
/protein_id="AAD12310.1"
/db_xref="PID:94219096"
/db_xref="GI:4219096"
/translation="MCTIFVQSRFCIFPLTFKSSASPKFLTNVTGCGSNTVTRPL
SSKVLTAVDRLRCP"
553..732
/codon_start=1
/product="ORF-12"
/protein_id="AAD12311.1"
/db_xref="PID:94219097"
/db_xref="GI:4219097"
/translation="MYTSLWHLGVAKANGLLILOTRKPHGTGNHLETSGGMVTWVKW
LLMTFMAGCRGMV"
complement(688..753)
/codon_start=1
/product="ORF-8"
/protein_id="AAD12312.1"
/db_xref="PID:94219098"
/db_xref="GI:4219098"
/translation="MDTHTVSDHPTAASHKSHQ"
983..988
complement(989..1033)
/codon_start=1
/product="ORF-11"
/protein_id="AAD12313.1"
/db_xref="PID:94219099"
/db_xref="GI:4219099"
/translation="MNNKNHYEVIKKTQ"
1016..1177
/codon_start=1
/product="ORF-5"
/protein_id="AAD12314.1"
/db_xref="PID:94219100"
/db_xref="GI:4219100"
/translation="MVFIHLGFKWGFKEKPSELYINGYTDIVVLVYVTFERSAEA
YMWYISSSL"
complement(1022..1027)
complement(1256..1735)
/note="similar to Porcine circovirus ORF-2 product encoded
by the sequence presented in GenBank Accession Number
AF027217"
/codon_start=1
/product="ORF-2"
/protein_id="AAD12315.1"
```

```
/db_xref="PID:94219101"  
/db_xref="GI:4219101"  
/translation="MTFPRRRYRRRRHPSRLGOLLRRPWFLLPHRRYRWRKNGI  
ENTLRSFGYTKATVTPPSWADMLRFKIDDFVPGSGTNKISLPEYVIRKVK  
VEFWPCSIITQGDROGVSSAVILDDNFFPKSIALYIDPYVNYSSRHTITQPSYHSR"  
1524..1631  
/codon_start=1  
/product="ORF-10"  
/protein_id="AAD12317.1"  
/db_xref="PID:94219103"  
/db_xref="GI:4219103"  
/translation="MSTAQEGVLTVVLTVPKVRERRVLKMPFFLLQR"  
complement(1528..1611)  
/codon_start=1  
/product="ORF-6"  
/protein_id="AAD12316.1"  
/db_xref="PID:94219102"  
/db_xref="GI:4219102"  
/translation="WASSTPASPPSDILSRLPQSERPPGR"  
1682..1741  
/codon_start=1  
/product="ORF-7"  
/protein_id="AAD12318.1"  
/db_xref="PID:94219104"  
/db_xref="GI:4219104"  
/translation="MAAGAVSSSAVTPPWKRHS"  
join(1750..1768,1..13)  
/note="putative replication site"  
join(1762..1768,1..2)  
/note="similar to the nonnucleotide motif of Porcine  
circovirus presented in GenBank Accession Numbers AF027217  
and U49186."
```

CDS

CDS

CDS

stem\_loop

rep\_origin

```
BASE COUNT 452 a 358 c 497 g 461 t  
ORIGIN
```

```
alignment_scores:  
  Quality: 1674.00      Length: 314  
  Ratio: 5.365          Gaps: 0  
Percent Similarity: 99.363 Percent Identity: 96.815
```

```
alignment_block:  
US-09-209-961-3 x AF117753 ..
```

```
Align seg 1/1 to: AF117753 from: 1 to: 1768
```

```
1 MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTr 17  
|||||  
51 ATGCCAGCAAAAAGATGGAAGAAGCGGCCCAACACACAAAAGGTG 100  
17 pValPheThrLeuAsnAsnProSerGluAspGluArgLysLysIleArg 34  
|||||  
101 GGTGTTCCAGCTGAATAATCCTCCGAAGACGAGCGCAAGAAATACGGG 150  
34 LuLeuProIleSerLeuPheAspTyrPheIleValGlyGluGluGlyAsn 50  
|||||  
151 AGCTTCCAAATCTCCCTTTTGAATATTTTATTTGTTGGCAGAGGGTAAT 200  
51 GluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPheValLysLy 67  
|||||  
201 GAGGAAGGACGACACCCACCCTCCAGGGTTCGCTAATTTTGTGAAGAA 250  
67 sGlnThrPheAsnLysValLysTrpTyrLeuGlyAlaArgCysHisIleG 84  
|||||  
251 GCAAACATTTAATAAAGTGAATGATTTTGGTGCCCGCTGCCACATCG 300  
84 LuLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrCysSerLysGlu 100  
|||||  
301 AGAAAGCCAAAGGAACTGATCAGCAGATAAAGAAATATTGCAAGTAAAGAA 350  
101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArgse 117  
|||||  
351 GGCACACTTACTGATGGAAATGTGGAGCTCCTAGATCTCAAGGACACCGGAG 400
```



Correct: 0.15553 + 0.0000

Sequence	Strd Orig	zScore	EScore	Len	Documentation
N_Geneseq_36:X35213	+	1882.00	3194.43	2,5e-170	1768
N_Geneseq_36:X35212	+	1682.00	3194.43	2,5e-170	1768
N_Geneseq_36:X35210	+	1676.00	3182.93	1,1e-169	1767
N_Geneseq_36:X35211	+	1676.00	3182.93	1,1e-169	1767
N_Geneseq_36:X35012	+	1640.00	3113.89	7,6e-166	1768
N_Geneseq_36:X35014	+	1640.00	3113.89	7,6e-166	1768
N_Geneseq_36:X35013	-	1447.00	2743.88	3,1e-145	1759
N_Geneseq_36:T13161	+	176.00	312.80	8,0e-10	1022
N_Geneseq_36:T171832	+	168.50	297.54	5,7e-09	1106
N_Geneseq_36:T171833	+	163.00	287.09	2,2e-08	1096
N_Geneseq_36:T171834	+	156.00	273.72	1,2e-07	1091
N_Geneseq_36:V24093	+	156.00	273.58	1,2e-07	1105
N_Geneseq_36:V24091	+	152.00	265.93	3,3e-07	1103
N_Geneseq_36:V24092	+	152.00	265.92	3,3e-07	1104
N_Geneseq_36:V24077	+	147.50	257.22	1,0e-06	1111
N_Geneseq_36:V24084	+	147.50	257.22	1,0e-06	1111
N_Geneseq_36:V24085	+	147.50	257.22	1,0e-06	1111
N_Geneseq_36:V24086	+	144.50	251.48	2,1e-06	1110
N_Geneseq_36:V24090	+	144.50	251.47	2,1e-06	1111
N_Geneseq_36:V24088	+	143.50	249.57	2,7e-06	1109
N_Geneseq_36:V24087	+	143.50	249.55	2,7e-06	1111
N_Geneseq_36:V24089	+	140.50	243.81	5,6e-06	1110
N_Geneseq_36:T19405	+	124.00	213.53	0,0003	982
N_Geneseq_36:Q12023	+	132.50	206.88	0,0006	7722
N_Geneseq_36:Q56826	+	132.50	206.84	0,0006	7753
N_Geneseq_36:Q03418	+	114.00	172.34	0,0534	7100
N_Geneseq_36:Q03360	+	114.00	172.34	0,0534	7100
N_Geneseq_36:N81393	+	114.00	172.34	0,0534	7102
N_Geneseq_36:N60044	+	114.00	172.34	0,0534	7102
N_Geneseq_36:Q34687	+	99.50	167.70	0,0969	886
N_Geneseq_36:T193024	+	111.00	166.13	0,1185	7400
N_Geneseq_36:N81394	+	103.00	162.91	0,1790	2490
N_Geneseq_36:T13165	+	95.00	157.54	0,3566	1017
N_Geneseq_36:N80183	+	103.00	151.17	0,8071	7152
N_Geneseq_36:N81390	+	103.00	151.17	0,8071	7152
N_Geneseq_36:T193023	+	102.50	149.69	0,9755	7493
N_Geneseq_36:N60080	+	102.50	149.69	0,9757	7494
N_Geneseq_36:Q03389	+	105.00	149.40	1,01	11832
N_Geneseq_36:Q05889	+	102.00	148.73	1,10	7494
N_Geneseq_36:Q03512	+	99.00	142.98	2,31	7493
N_Geneseq_36:T19331	+	87.00	140.88	3,02	1145
N_Geneseq_36:T135035	+	86.00	140.81	3,38	1042
N_Geneseq_36:X35213	+	1882.00	3194.43	2,5e-170	1768
N_Geneseq_36:X35212	+	1682.00	3194.43	2,5e-170	1768
N_Geneseq_36:X35210	+	1676.00	3182.93	1,1e-169	1767
N_Geneseq_36:X35211	+	1676.00	3182.93	1,1e-169	1767
N_Geneseq_36:X35012	+	1640.00	3113.89	7,6e-166	1768
N_Geneseq_36:X35014	+	1640.00	3113.89	7,6e-166	1768
N_Geneseq_36:X35013	-	1447.00	2743.88	3,1e-145	1759
N_Geneseq_36:T13161	+	176.00	312.80	8,0e-10	1022
N_Geneseq_36:T171832	+	168.50	297.54	5,7e-09	1106
N_Geneseq_36:T171833	+	163.00	287.09	2,2e-08	1096
N_Geneseq_36:T171834	+	156.00	273.72	1,2e-07	1091
N_Geneseq_36:V24093	+	156.00	273.58	1,2e-07	1105
N_Geneseq_36:V24091	+	152.00	265.93	3,3e-07	1103
N_Geneseq_36:V24092	+	152.00	265.92	3,3e-07	1104
N_Geneseq_36:V24077	+	147.50	257.22	1,0e-06	1111
N_Geneseq_36:V24084	+	147.50	257.22	1,0e-06	1111
N_Geneseq_3					

.....

598 GCAAACTTTTAAATAAAGTGAAGTGGTATTGGGTGCCCGCTGCCACATCG 64



```

|||||
648 AGAAGCCAAAGGAAGTGTATCAGCAAGTAATAAGATATTCAGTAAGAA 697
|||||
101 GlyAsnLeuLeuLeuGluCysGlyAlaProArgSerGlnGlnArgSe 117
|||||
698 GGCACCTTACTTATTAAGTGTGAGCTCCCTCGATCTCAAGGACAACGGAG 747
|||||
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValt 134
|||||
748 TGACCTGTCTACTGTGTGAGTACCTTGTGGAGAGCGGAGTGTGTGA 797
|||||
134 hrValAlaGluGlnHisProValThrPheValLysAsnPheArgGlyLeu 150
|||||
798 CCGTTCAGACGACGACCCCTGTAACTGTTTCAGAAATTCGCGGGCTG 847
|||||
151 AlaGluLeuLeuLysValSerGlyLysMetGlnLysArgAspTrpLysTh 167
|||||
848 GCTGAAGCTTTTGAAGTGTAGCGGGAATAATCAGAAGCGTGTATGGAAGAC 897
|||||
167 rAsnValHisPheIleValGlyProGlyCysGlyLysSerLysTrpA 184
|||||
898 CAATGTACAGCTATTGTGGGCGCCACCTGGTGTGTAAGCAATGGG 947
|||||
184 laAlaAsnPheAlaAsnProGluThrThrTrpTrpLysProProLysAsn 200
|||||
948 CTGCTAATTTTGCAGACCCCGGAACACATACCTAGAAACCCACCTAGA 997
|||||
201 LysTrpTrpAspGlyTrpHisGlyGluLysValValValIleAspAspPh 217
|||||
998 AAGTGTGGGATGGTACCATGTGTGAAGAGTGTGTTTATTTGATGACTT 1047
|||||
217 eTrpGlyTrpLeuProTrpAspLeuLeuArgLeuLysAspArgTrpP 234
|||||
1048 TTATGCTGGCTCCCGTGGATGATCTACTGAGACTGTGTGATCGATATC 1097
|||||
234 roLeuThrValLysThrLysGlyGlyThrValProPheLeuAlaArgSer 250
|||||
1098 CATTGACTGTAGAGACTAAAGTGGAACTGTACCTTTTGGCCCGCAGT 1147
|||||
251 IleLeuIleThrSerAsnGlnThrProLeuGluTrpTrpSerSerThrAl 267
|||||
1148 ATTCTGATTACCAAGCAATCAGACCCCGTGGTAATGGTACTCCCACTGC 1197
|||||
267 aValProAlaValGluAlaLeuTrpArgArgIleThrSerLeuValPheT 284
|||||
1198 TGTCCAGCTGTAGAGCTCTCTATCGGAGGATTAATCTCTTGGTATTTT 1247
|||||
284 rPlysAsnAlaThrLysGlnSerThrGluGluGlyGlnPheValThr 300
|||||
1248 GGAAGAATGCTACAGAACAATCCACGGAGGAGGGGCCAGTTCGTCAACC 1297
|||||
301 LeuSerProProCysProGluPheProTrpGluIleAsnTrp 314
|||||
1298 CTTTCCCCCATCGCTGAATTTCCATATGAATAAATATAC 1339

```

seq\_name: N\_Geneseq\_36.X35212

seq\_documentation\_block:

```

ID X35212 standard; DNA; 1768 BP.
AC X35212;
DT 01-JUL-1999 (first entry)
DE Nucleotide sequence of PCV isolate Imp999 (corrected version).
KW PCV isolate; type II porcine circovirus; PCV; PMWS;
KW porcine multisystemic wasting syndrome; pig; vaccine; ss.
OS Porcine circovirus.
PN FR2769322-Al.
PD 09-APR-1999.
PE 22-JAN-1998; 000873.
PR 03-OCT-1997; FR-012382.
PA (MERI-) MERIAL SAS.
PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,
PI Haines D, Harding J, Hassard L, Meehan B;
DR WPI; 99-246948/21.

```

PT New type II porcine circovirus, used for, e.g. passive immunization  
 PT of pregnant sows  
 PS Claim 14; Fig 3: 48pp; French.  
 CC The present sequence represents the nucleotide sequence of PCV isolate  
 CC Imp999. The specification describes a preparation of type II  
 CC porcine circovirus (PCV), which is particularly isolated from a lesion,  
 CC from a pig with symptoms of PMWS (porcine multisystemic wasting  
 CC syndrome). PCV (attenuated or inactivated), polypeptides derived from  
 CC it, and vectors that express these polypeptides are all useful in  
 CC vaccines, suitable for administration to adult or young pigs, or to  
 CC pregnant sows (for passive immunization of their offspring). DNA  
 CC isolated from PCV is used for in vivo or in vitro expression of viral  
 CC polypeptides, also as probes or primers for diagnosis in usual  
 CC hybridization or amplification assays. These polypeptides may also be  
 CC used diagnostically to detect PCV-specific antibodies, while antibodies  
 CC raised against the polypeptides can be used to detect antigens, in any  
 CC usual immunoassay format.  
 CC  
 SQ Sequence 1768 BP; 453 A; 360 C; 494 G; 461 T;

alignment\_scores:

Quality: 1682.00 Length: 314  
 Ratio: 5.391 Gaps: 0  
 Percent Similarity: 99.363 Percent Identity: 97.452

alignment\_block:

US-09-209-961-3 x X35212

Align seg 1/1 to: X35212 from: 1 to: 1768

```

1 MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTr 17
|||||
398 ATGCCACGACGAGAAATGGAAGAGCGGACCCCAACACACATAAAGGTG 447
|||||
17 pValPheThrLeuAsnAsnProSerGluAspGluArgLysLysIleArgG 34
|||||
448 GGTGTTCCGCTGAATATCTCTCCGAGAGCGGCGCAAGAAATACGGG 497
|||||
34 LuLeuProIleSerLeuPheAspTrpPheIleValGlyGluGluGlyAsn 50
|||||
498 AGCTCCCAATCTCCCTATTTATTTATTTATTTGTCGAGAGGGGTAA 547
|||||
51 GluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPheValLysLy 67
|||||
548 GAGGAGGACGACACCTCACCTCCAGGGGTTCGCTAATTTGTGAAGAA 597
|||||
67 sGlnThrPheAsnLysValLysTrpTrpLeuGlyAlaArgCysHisIleG 84
|||||
598 GCAAACTTTTAAAGTGAAGTGTATTTGGTGCCGCTGCCACATCG 647
|||||
84 LuLysAlaLysGlyThrAspGlnGlnAsnLysGluTrpCysSerLysGlu 100
|||||
648 AGAAGGCCAAAGGAACTGATCAGCAGAATAAGAAATATTGCAGTAAAGAA 697
|||||
101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArgSe 117
|||||
598 GGCACCTTACTTATTAAGTGTGAGTACCTCTCGATCTCAAGGACAACGGAG 747
|||||
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValt 134
|||||
748 TGACCTGTCTACTGTGTGAGTACCTCTCTGGAGAGCGGAGTGTGTGA 797
|||||
134 hrValAlaGluGlnHisProValThrPheValLysAsnPheArgGlyLeu 150
|||||
798 CCGTTCAGACGACGACCCCTGTAACTGTTTCAGAAATTCGCGGGCTG 847
|||||
151 AlaGluLeuLeuLysValSerGlyLysMetGlnLysArgAspTrpLysTh 167
|||||
848 GCTGAAGCTTTTGAAGTGTAGCGGGAATAATCAGAAGCGTGTATGGAAGAC 897
|||||
167 rAsnValHisPheIleValGlyProGlyCysGlyLysSerLysTrpA 184
|||||
948 CTGCTAATTTTGCAGACCCCGGAACACATACCTAGAAACCCACCTAGA 997
|||||
201 LysTrpTrpAspGlyTrpHisGlyGluLysValValValIleAspAspPh 217
|||||
998 AAGTGTGGGATGGTACCATGTGTGAAGAGTGTGTTTATTTGATGACTT 1047
|||||
217 eTrpGlyTrpLeuProTrpAspLeuLeuArgLeuLysAspArgTrpP 234
|||||
1048 TTATGCTGGCTCCCGTGGATGATCTACTGAGACTGTGTGATCGATATC 1097
|||||
234 roLeuThrValLysThrLysGlyGlyThrValProPheLeuAlaArgSer 250
|||||
1098 CATTGACTGTAGAGACTAAAGTGGAACTGTACCTTTTGGCCCGCAGT 1147
|||||
251 IleLeuIleThrSerAsnGlnThrProLeuGluTrpTrpSerSerThrAl 267
|||||
1148 ATTCTGATTACCAAGCAATCAGACCCCGTGGTAATGGTACTCCCACTGC 1197
|||||
267 aValProAlaValGluAlaLeuTrpArgArgIleThrSerLeuValPheT 284
|||||
1198 TGTCCAGCTGTAGAGCTCTCTATCGGAGGATTAATCTCTTGGTATTTT 1247
|||||
284 rPlysAsnAlaThrLysGlnSerThrGluGluGlyGlnPheValThr 300
|||||
1248 GGAAGAATGCTACAGAACAATCCACGGAGGAGGGGCCAGTTCGTCAACC 1297
|||||
301 LeuSerProProCysProGluPheProTrpGluIleAsnTrp 314
|||||
1298 CTTTCCCCCATCGCTGAATTTCCATATGAATAAATATAC 1339

```

184 laAlaAsnPhleAlaAsnProGluThrThrTyrTrpLysProProLysAsn 200  
|||||  
948 CTGCTAATTTTGCAGACCGGAACCATACACTGGAAACACCTAGAAAC 997  
|||||  
201 LysTrpTrpAspGlyTyrHisGlyGluLysValValValIleAspAsp 217  
|||||  
998 AAGTGGTGGATGTTACCATGGTGAAGAACTGGTGTATTGATGACTT 1047  
|||||  
217 eTyrGlyTrpLeuProTrpAspAspLeuLeuArgLeuLysAspArgTyr 234  
|||||  
1048 TTATGGCTGGCTGGCGGGATGATCTACTGAGACTGTGTGATCATATC 1097  
|||||  
234 roLeuThrValLysThrLysGlyGlyThrValProPheLeuAlaArgSer 250  
|||||  
1098 CATGACTGTAGACATAAGCTGAAGTGAAGTGTACCTTTTGGCCCGCACT 1147  
|||||  
251 IleLeuIleThrSerAsnGlnThrProLeuGluTrpTyrSerSerThrAl 267  
|||||  
1148 ATTCGTATTACCAGCATCAGACCCGTTGGAATGGTACTCCTCAACTGC 1197  
|||||  
267 aValProAlaValGluAlaLeuTyrArgArgIleThrSerLeuValPheT 284  
|||||  
1198 TGTCCAGCTAGAGCTCTCTATCGGAGGATTACTTCTCTGGTATTAT 1247  
|||||  
284 rPLysAsnAlaThrLysGlnSerThrGluGluGlyGlyGlnPheValThr 300  
|||||  
1248 GGAAGATGCTACAGACAACTCCAGGAGGAGGGCCAGTTCTGTCACC 1297  
|||||  
301 LeuSerProProCysProGluPheProTyrGluIleAsnTyr 314  
|||||  
1298 CTTTCCCCCATCGCCCTGAATTTCCATATCAATAAATTAC 1339  
|||||

seq\_name: N\_Geneseq\_36.X35210

seq\_documentation\_block:

ID X35210 standard; DNA; 1767 BP.  
AC X35210;  
DT 01-JUL-1999 (first entry)  
DE Nucleotide sequence of PCV isolate Impl011-48121.  
KW PCV isolate; type II porcine circovirus; PCV; PMWS;  
OS porcine multisystemic wasting syndrome; pig; vaccine; ss.  
KS Porcine circovirus.  
PN FR2769322-Al.  
PD 09-APR-1999.  
PF 22-JAN-1998; 000873.  
PR 03-OCT-1997; FR-012382.  
PA (MERI-) MERIAL SAS.  
PI Allan G. Chappuis GE, Charreyre CE, Clark E, Ellis J,  
PI Haines D, Harding J, Hassard L, Meehan B;  
DR WPI; 99-246948/21.  
PT New type II porcine circovirus, used for, e.g. passive immunization  
PT of pregnant sows  
PS Claim 14: fig 1: 48pp; French.  
CC The present sequence represents the nucleotide sequence of PCV isolate  
CC Impl011-48121. The specification describes a preparation of type II  
CC porcine circovirus (PCV), which is particularly isolated from a lesion,  
CC from a pig with symptoms of PMWS (porcine multisystemic wasting  
CC syndrome). PCV (attenuated or inactivated), polypeptides derived from  
CC it, and vectors that express these polypeptides are all useful in  
CC vaccines, suitable for administration to adult or young pigs, or to  
CC pregnant sows (for passive immunization of their offspring). DNA  
CC isolated from PCV is used for in vivo or in vitro expression of viral  
CC polypeptides, also as probes or primers for diagnosis in usual  
CC hybridization or amplification assays. These polypeptides may also be  
CC used diagnostically to detect PCV-specific antibodies, while antibodies  
CC raised against the polypeptides can be used to detect antigens, in any  
CC usual immunoassay format.  
SQ Sequence 1767 BP; 447 A; 360 C; 502 G; 458 T;

alignment\_scores:

Quality: 1676.00 Length: 314

Ratio: 5.372 Gaps: 0  
Percent Similarity: 99.363 Percent Identity: 96.815  
alignment\_block:  
US-09-209-961-3 x X35210 ..  
Align seg 1/1 to: X35210 from: 1 to: 1767  
1 MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTr 17  
|||||  
398 ATGCCGAGCAAGAAGTGAAGAAGCGGACCCCAACCCCATAAAGGTG 447  
|||||  
17 pValPheThrLeuAsnAsnProSerGluArgLysLysLysLysLysArg 34  
|||||  
448 GGTGTCTACTCTGAATAATCCCTCCGAGACGAGCGCAAGAAATACGGG 497  
|||||  
34 LuLeuProIleSerLeuPheAspTyrPheIleValGlyGluGluGlyAsn 50  
|||||  
498 ATCTTCCAATATCCCTATTATTGATTATTTATTGTCGAGGAGGTAAT 547  
|||||  
51 GluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPheValLysLy 67  
|||||  
548 GAGGAAGGACGACACCTCACCCTCCAGGGGTTCGCTAATTTTGTGAAGAA 597  
|||||  
67 sGlnThrPheAsnLysValLysTyrTyrLeuGlyAlaArgCysHisIleG 84  
|||||  
598 GCAGACTTTTAAAGTGAAGTGAAGTATTTGGTGCCCGCTGCCACATCG 647  
|||||  
84 LuLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrCysSerLysGlu 100  
|||||  
648 AGAAAGCGAAAGGACAGATCAGCAGATAAAGATAATGTCAGTAAAGAA 697  
|||||  
101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArgSe 117  
|||||  
698 GGCAACTTACTGATGGAGTGTGGAGCTCTAGATCTCAGGGACAACGGAG 747  
|||||  
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValT 134  
|||||  
748 TGACCTGTCTACTGCTGTGAGTACCTTTTGGAGCGGGGAGTCTGGTGA 797  
|||||  
134 hrValAlaGluGlnHisProValThrPheValLysAsnPheArgGlyLeu 150  
|||||  
798 CGTTGTCAGAGCAGCACCTCTGTAACTGTTGTCAGAAATTTCCCGGGCTG 847  
|||||  
151 AlaGluLeuLysValSerGlyLysMetGlnLysArgAspTrpLysTh 167  
|||||  
848 GCTGAACCTTTTGAAGTGAAGCGGAAATGCAGAACGCTGATTGGAAGAC 897  
|||||  
167 rAsnValHisPheIleValGlyProProGlyCysGlyLysSerLysTrpA 184  
|||||  
898 TAATGTACACGCTATTGTGGGGCCACCTGGGTGGTAAAGCAAAATGGG 947  
|||||  
184 laAlaAsnPhleAlaAsnProGluThrThrTyrTrpLysProProLysAsn 200  
|||||  
948 CTGCTAATTTTGCAGACCGGAACCATACACTGGAAACACCTAGAAAC 997  
|||||  
201 LysTrpTrpAspGlyTyrHisGlyGluLysValValValIleAspAsp 217  
|||||  
998 AAGTGGTGGATGTTTACCATGGTGAAGAACTGGTGTATTGATGACTT 1047  
|||||  
217 eTyrGlyTrpLeuProTrpAspAspLeuLeuArgLeuLysAspArgTyrP 234  
|||||  
1048 TTATGGCTGGCTGGCCCTGGGATGATCTACTGAGACTGTGTGATCATATC 1097  
|||||  
234 roLeuThrValLysThrLysGlyGlyThrValProPheLeuAlaArgSer 250  
|||||  
1098 CATGACTGTAGACATAAGCTGAAGTGAAGTGTACCTTTTGGCCCGCACT 1147  
|||||  
251 IleLeuIleThrSerAsnGlnThrProLeuGluTrpTyrSerSerThrAl 267  
|||||  
1148 ATTCGTATTACCAGCAATCAGACCCGTTGGAATGGTACTCCTCAACTGC 1197  
|||||  
267 aValProAlaValGluAlaLeuTyrArgArgIleThrSerLeuValPheT 284

```

1198 TGCCAGCTGTAGAACTCTTTATCGAGGATTACTTCCTGTTATTT 1247
284 rPLysAsnAlaThrLysGlnSerThrGluGluGlyGlnPheValThr 300
1248 GGAAGAAATGCTACAGAACAAATCCACGGAGGAGGCGCCAGTTCGTACC 1297
301 LeuSerProCysProGluPheProTyrGluIleAsnTyr 314
1298 CTTTCCCCCATGCCCTGAATTCATATGAATAAATTAC 1339

```

seq\_name: N\_Geneseq\_36:X35211

seq\_documentation\_block:

```

ID X35211 standard; DNA; 1767 BP.
AC X35211.
DT 01-JUL-1999 (first entry)
DE Nucleotide sequence of PCV isolate Impl011-48285.
KW PCV isolate; type II porcine circovirus; PCV; PMWS;
OS Porcine multisystemic wasting syndrome; pig; vaccine; ss.
PN FR2769322-Al.
PD 09-APR-1999.
PF 22-JAN-1998; 000873.
PR 03-OCT-1997; FR-012382.
PA (MERI-) MERIAL SAS.
PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,
PI Haines D, Harding J, Hassard L, Meehan B;
DR WPI; 99-246948/21.
PT New type II porcine circovirus, used for, e.g. passive immunization
PT of pregnant sows
PS Claim 14; Fig 2; 48pp; French.
CC The present sequence represents the nucleotide sequence of PCV isolate
CC Impl011-48285. The specification describes a preparation of type II
CC porcine circovirus (PCV), which is particularly isolated from a lesion,
CC from a pig with symptoms of PMWS (porcine multisystemic wasting
CC syndrome). PCV (attenuated or inactivated), polypeptides derived from
CC it, and vectors that express these polypeptides are all useful in
CC vaccines, suitable for administration to adult or young pigs, or to
CC pregnant sows (for passive immunization of their offspring). DNA
CC isolated from PCV is used for in vivo or in vitro expression of viral
CC polypeptides, also as probes or primers for diagnosis in usual
CC hybridization or amplification assays. These polypeptides may also be
CC used diagnostically to detect PCV-specific antibodies, while antibodies
CC raised against the polypeptides can be used to detect antigens, in any
CC usual immunoassay format.
SQ Sequence 1767 BP; 448 A; 359 C; 500 G; 460 T;

```

alignment\_scores:

Quality	1676.00	Length	314
Ratio	5.372	Gaps	0
Percent Similarity	99.363	Percent Identity	96.815

alignment\_block:

US-09-209-961-3 x X35211

Align seg 1/1 to: X35211 from: 1 to: 1767

```

1 MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTr 17
398 ATGCCAGCAAGAAGATGCAAGAGCGGACCCCAACCCCAATAAAGTG 447
17 pValPheThrLeuAsnAsnProSerGluAspGluArgLysLysIleArg 34
448 GGTGTTCACTCTGAATAATCTCCGAAGACGAGCGCAAGAAATACGG 497
34 luleuProIleSerLeuPheAspTyrPheIleValGlyGluGlyAsn 50
498 ATCTTCAATATCCCTATTTGATTTATTTGTCGAGGAGGGTAAT 547
51 GluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPheValLys 67

```

```

548 GAGGAGGACGAACACCTCACCTCCAGGGTTCGCTAATTTTGTGAAGAA 597
67 sGlnThrPheAsnLysValLysTyrTrpTyrLeuGlyAlaAArgCysHisIleG 84
598 GCAGACTTTTAAAGTGAAGTGGTATTTGGGTGCCCGCTGCCACATCG 647
84 lUlysAlaLysGlyThrAspGlnGlnAsnLysGluTyrCysSerLysGlu 100
648 AGAAAGCGAAGGAACAGATCAGCAGAGATAAAGATACTGCAGTAAAGAA 697
101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArgse 117
698 GGCAACTTACTGATGGAGTGTGGAGCTCTAGATCTCAGGCAACACGGAG 747
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValT 134
748 TGACCTGTCTACTGCTGTGAGTACTTGTGGAGAGCGGAGTCTGGTGA 797
134 hrValAlaGluGlnHisProValThrPheValLysAsnPheArgGlyLeu 150
798 CCGTTGCAGACGACACCTCTGAAGTTTGTCAAGAAATTCGCGCGGCTG 847
151 AlaGluLeuLeuLysValSerGlyLysMetGlnLysArgAspTrpLysTh 167
848 GCTGAACCTTTTGAAGTGAAGCGGGAATGCAGAGCGGTGATTGGAAGAC 897
167 rAsnValHisPheIleValGlyProGlyCysGlyLysSerLysTyrPA 184
898 TAAITGACACTATTGTGGGCGCACCTGGGTGGTAAAGCAAAATGGG 947
184 laAlaAsnPheAlaAsnProGluThrThrTyrTrpLysProLysAsn 200
948 CTGCTAATTTTGCAGACCCGGAACCACTACTGGAACACCTAGAAAC 997
201 LysTrpIleAspGlyTyrHisGlyGluLysValValIleAspAspPh 217
998 AAGTGGTGGGATGTTTACCATGGTGAAGAGTGGTGTATTATGATGACTT 1047
217 eTyrGlyTrpLeuProTyrAspAspLeuArgLeuCysAspArgTyrP 234
1048 TTAITGGCTGGCTGGCCCTGGGATGATCTACTGAGACTGTGTGATCATATC 1097
234 rOleuThrValLysThrLysGlyThrValProPheLeuAlaArgSer 250
1098 CATTGACTGTAGAGACTTAAAGTGAAGTGTACCTTTTGGCCCGCAGT 1147
251 IleLeuIleThrSerAsnGlnThrProLeuGluTyrTrpTyrSerThrAl 267
1148 ATTCTGATTACCAGCAATCAGACCCCGTTGGAATGGTACTCTCACTCACTGC 1197
267 aValProAlaValGluAlaLeuTyrArgArgIleThrSerLeuValPheP 284
1198 TGTCCAGCTGTAGAAGCTCTTTATCGAGGAGTACTTCTCTGGTATTTT 1247
284 rPLysAsnAlaThrLysGlnSerThrGluGluGlyGlnPheValThr 300
1248 GGAAGATGCTACAGAACAAATCCACGGAGGAGGCGCCAGTTCGTACC 1297
301 LeuSerProCysProGluPheProTyrGluIleAsnTyr 314
1298 CTTTCCCCCATGCCCTGAATTTCCATATGAATAAATTAC 1339

```

seq\_name: N\_Geneseq\_36:X35012

seq\_documentation\_block:

```

ID X35012 standard; DNA; 1768 BP.
AC X35012.
DT 01-JUL-1999 (first entry)
DE Genomic DNA sequence of PCV strain 999PCV.
KW PCV; strain PCV999; pig; PMWS; porcine multisystemic wasting syndrome;
OS Porcine circovirus
PN FR2769321-Al.

```

PD 09-APR-1999.  
PF 03-OCT-1997: 012382.  
PR 03-OCT-1997: FR-012382.  
PA (MERI-) MERIAL SAS.  
PA (UYBE-) UNIV QUEENS BELFAST.  
PA (UYSA-) UNIV SASKATCHEWAN.  
PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,  
PI Haines D, Harding J, Hassard L, Meehan B;  
DR WPI: 99-246947/21.  
PT New porcine circovirus from animals with porcine systemic wasting  
PT Syndrome 13;  
PS Claim 13; Fig 1: 35pp; French.  
CC The present sequence represents the genomic sequence of porcine  
CC circovirus (PCV) strain PCV999. The sequence is obtained from PCV  
CC isolated from a physiological or tissue sample, particularly from  
CC a lesion, from a pig showing symptoms of PMWS (porcine multisystemic  
CC wasting syndrome), or cultured cells, infected with PCV isolated from  
CC such pigs. PCV (attenuated or inactivated), polypeptides derived from  
CC it, and vectors that express these polypeptides are all useful in  
CC vaccines, suitable for administration to adult or young pigs, or to  
CC pregnant sows (for passive immunization of their offspring). DNA  
CC isolated from PCV is used for in vivo or in vitro expression of viral  
CC polypeptides, also as probes or primers for diagnosis in usual  
CC hybridization or amplification assays. These polypeptides may also be  
CC used diagnostically to detect PCV-specific antibodies, while antibodies  
CC raised against the polypeptides can be used to detect antigens, in any  
CC usual immunoassay format.  
SQ Sequence 1768 BP; 453 A; 359 C; 486 G; 455 T;

alignment\_scores:  
Quality: 1640.00 Length: 314  
Ratio: 5.359 Gaps: 0  
Percent Similarity: 97.452 Percent Identity: 95.541

alignment\_block:  
US-09-209-961-3 x X35012

Align seg 1/1 to: X35012 from: 1 to: 1768

1 MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTr 17  
|||||  
399 ATGCCCCAGCAAGAAGTGGAGAGCGGACCCCAACACATAAAGGTG 448  
|||||  
17 pValPheThrLeuAsnAsnProSerGluAspGluArgLysLysIleArgG 34  
|||||  
449 GGTGTTCAGCTGAATAATCTCTCCAGACGACGCGCAAGAAATACGGG 498  
|||||  
34 LuLeuProIleSerLeuPheAspTyrPheIleValGlyGluGlyAsn 50  
|||||  
499 AGCTCCCAATCTCCCTATTGTAATTTATTTATTTGTTGGCAGAGGGTWT 548  
|||||  
51 GluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPheValLysLy 67  
|||||  
549 GAGGAAGACGACGACCTCACCTCCAGGGGTTCGCTAATTTTGTGAAGAA 598  
|||||  
67 sGlnThrPheAsnLysValLysTrpTyrLeuGlyAlaArgCysHisIleG 84  
|||||  
599 GCAAACTTTTAATAAGTCAAGTGTGTAATTTGTTGGTCCCGCTCCACATCG 648  
|||||  
84 LuLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrCysSerLysGlu 100  
|||||  
649 AGAAAGCCAAAGAACTGATCATCAGCAGATAAAGATAATTTCAGTAAAGAA 698  
|||||  
101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArgSe 117  
|||||  
699 GGCAACTTACTTATTGAATGTGGAGCTCTCGATCTCAAGGACAAACGGAG 748  
|||||  
117 AspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValT 134  
|||||  
749 TGACCTGTCTACTGCTGTGAGTACCTGTTGGAGAGCGGGAGTCTGGTGA 798  
|||||  
134 hrValAlaGluGlnHisProValThrPheValLysAsnPheArgGlyLeu 150

799 CCGTTGCAGAGCAGCACCTGTAACGTTTGTGAGAAATTTCCGCGGCTG 848  
|||||  
151 AlaGluLeuLeuLysValSerGlyLysMetGlnLysArgAspTrpLysTh 167  
|||||  
849 GCTGAACCTTTTGAAGTGAAGCGGGAATAATCAGAAGCGTATTGGAAGAC 898  
|||||  
167 rAsnValHisPheIleValGlyProGlyCysGlyLysSerLysTrpA 184  
|||||  
899 CAATGTACACGTCATTGTGGGCCACCTGGGTGTGGTAAAGCAAAATGGG 948  
|||||  
184 laAlaAsnPheAlaAsnProGluThrThrTyrTrpLysProLysAsn 200  
|||||  
949 CTGCTAATTTTGCAGACCCGGAACACACATCTAGAAACCCACCTAGAAC 998  
|||||  
201 LysTrpTrpAspGlyTyrHisGlyGluLysValValIleAspAspPh 217  
|||||  
999 AGTGTGGGATGTTTACCATGTGTGAAGAGTGTGTTATTGATGACTT 1048  
|||||  
217 eTyrGlyTrpLeuProTrpAspAspLeuLeuArgLeuLysAspArgTyrP 234  
|||||  
1049 TTAATGGCTGCTCCCTGGGATGATCTCTAGAGACTGTGTGATCGATATC 1098  
|||||  
234 rLeuThrValLysThrLysGlyGlyThrValProPheLeuAlaArgSer 250  
|||||  
1099 CATGTACTGTAGAGACTAAGGTGGAACCTGTACNNNNNNNGGCCCGCAGT 1148  
|||||  
251 IleLeuIleThrSerAsnGlnThrProLeuGluTrpTyrSerSerThrAl 267  
|||||  
1149 ATTCTGATTACCAGCAATCAGACCCGCTTGGAAATGGTACTCCTCAACTGC 1198  
|||||  
267 aValProAlaValGluAlaLeuTyrArgArgIleThrSerLeuValPheT 284  
|||||  
1199 TGTCCAGCTGTAGAAGCTCTCTATCGAGGATTAATTAATTAATAC 1248  
|||||  
284 rLysAsnAlaThrLysGlnSerThrGluGlyGlyGlnPheValThr 300  
|||||  
1249 GGAGAAATGCTACAGAACATCCACGGAGAGGGGGCCAGTNGTCACC 1298  
|||||  
301 LeuSerProProCysProGluPheProTyrGluIleAsnTyr 314  
|||||  
1299 CTTTCCCCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1340

seq\_name: N\_Geneseq\_36:X35214

seq\_documentation\_block:  
ID X35214 standard; DNA; 1768 BP.  
AC X35214;  
DT 01-JUL-1999 (first entry)  
DE Nucleotide sequence of PCV isolate Imp999 (uncorrected version).  
KW PCV isolate; type II porcine circovirus; PCV; PMWS;  
KW porcine multisystemic wasting syndrome; pig; vaccine; ss.  
OS Porcine circovirus.  
PN FR2769322-A1.  
PD 09-APR-1999.  
PF 22-JAN-1998; 000873.  
PR 03-OCT-1997; FR-012382.  
PA (MERI-) MERIAL SAS.  
PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,  
PI Haines D, Harding J, Hassard L, Meehan B;  
DR WPI: 99-246948/21.  
PT New type II porcine circovirus, used for, e.g. passive immunization  
PT of pregnant sows  
PS Claim 14; Fig 6: 48pp; French.  
CC The present sequence represents the nucleotide sequence of PCV isolate  
CC Imp999. The specification describes a preparation of type II  
CC porcine circovirus (PCV), which is particularly isolated from a lesion,  
CC from a pig with symptoms of PMWS (porcine multisystemic wasting  
CC syndrome). PCV (attenuated or inactivated), polypeptides derived from  
CC it, and vectors that express these polypeptides are all useful in  
CC vaccines, suitable for administration to adult or young pigs, or to  
CC pregnant sows (for passive immunization of their offspring). DNA  
CC isolated from PCV is used for in vivo or in vitro expression of viral

CC polypeptides, also as probes or primers for diagnosis in usual  
CC hybridization or amplification assays. These polypeptides may also be  
CC used diagnostically to detect PCV-specific antibodies, while antibodies  
CC raised against the polypeptides can be used to detect antigens, in any  
CC usual immunoassay format.  
SQ Sequence 1768 BP; 453 A; 359 C; 486 G; 455 T;

alignment\_scores:  
Quality: 1640.00 Length: 314  
Ratio: 5.359 Gaps: 0  
Percent Similarity: 97.452 Percent Identity: 95.541

alignment\_block:

US-09-209-961-3 x X35214

Align seg 1/1 to: X35214 from: 1 to: 1768

1 MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTr 17  
|||||  
399 ATGCCCAAGCAAGAAATGGAAGAGCGGCCCAACACACATAAAGGTG 448  
17 pValPheThrLeuAsnAsnProSerGluAspGluArgLysLysLeuArg 34  
|||||  
449 GGTGTTCCAGCTCAATATCTCTCCGAGAGCGAGCGCAAGAAATACGGG 498  
34 luleProIleSerLeuPheAspTyrPheIleValGlyGluGluGlyAsn 50  
|||||  
499 AGCTCCCAATCTCCATTTGATTTATTTATTTGCGGAGGGGTTWT 548  
51 GluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPheValLysLy 67  
|||||  
549 GAGGAAGACGACACCTCACCTCCAGGGTTCCGTAATTTGTGAGAA 598  
67 sGlnThrPheAsnLysValLysTyrPheLeuGlyAlaAargCysHisLeG 84  
|||||  
599 GCAAACTTTTAATAAGTGAAGTGTATTTGGTGCCTGCGCATCG 648  
84 luleYAlaLysGlyThrAspGlnGlnAsnLysGluTyrCysSerLysGlu 100  
|||||  
649 AGAAGCCAAAGGAGATGATCAGCAGATAAAGATATTCGAGTAAGAA 698  
101 GlyAsnLeuLeuLeuGlyCysGlyAlaProArgSerGlnGlyGlnArgse 117  
|||||  
699 GGCAACTTATTGATGAGGAGCTTCGATCTCAAGGACACAGGAG 748  
117 AspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValt 134  
|||||  
749 TGACCTGCTACTGCTGAGTACCTTTGTTGGAGCGGGAGTCTGGTGA 798  
134 hrValAlaGluGlnHisProValThrPheValLysAsnPheArgGlyLeu 150  
|||||  
799 CCGTTCGAGACGACACCCCTGTACAGTTGTCAGAAATTTCCGCGGGGTG 848  
151 AlaGluLeuLysValSerGlyLysMetGlnLysArgAspTyrLysTh 167  
|||||  
849 GCTGAACCTTTTGAAGTACGGGGAATGCGAAGCGTGTATGGAAGAC 898  
167 rAsnValHisPheIleValGlyProProGlyCysGlyLysSerLysTrpa 184  
|||||  
899 CAATGTACAGTCAATCTGGGGCCACCTGGGTGTGTAAGCAAAATGGG 948  
184 laAlaAsnPheAlaAsnProGluThrThrTyrTrpLysProProLysAsn 200  
|||||  
949 CTGCTAATTTTGCAGACCCGGAACCACTACTGGAACACCACTAGAAAC 998  
201 LysTrpTrpAspGlyTyrHisGlyLysValValValIleAspAspPh 217  
|||||  
999 AAGTGGTGGGATGTTTACCATGGTGAAGAGAGTGTGTATGATGACTT 1048  
217 eTyrGlyTrpLeuProTrpAspAspLeuLeuArgLysCysAspArgTyrP 234  
|||||  
1049 TTATGGCTGGCTGCCGCGGAGTATCTACTGAGACTGTGTGATCATATC 1098

234 roLeuThrVallysThrLysGlyThrValProPheLeuAlaAargSer 250  
|||||  
1099 CATGTACTGTAGAGACTAAGGTGAAGTACTGACNNNNNNNGGCCGCGAGT 1148  
251 lIleLeuIleThrSerAsnGlnThrProLeuGluTyrPyrSerSerThrAl 267  
|||||  
1149 ATTCTGATTACCAAGCAATCAGACCCGTTGGAATGGTACTCCTCAACTGC 1198  
267 aValProAlaValGluAlaLeuTyrArgArgIleThrSerLeuValPheT 284  
|||||  
1199 TGTCCACGCTGTAGAAGCTCTCTATCGGAGGATTACTTCCTTGGTATTTT 1248  
284 rPLysAsnAlaThrLysGlnSerThrGluGlyGlyGlnPheValThr 300  
|||||  
1249 GGAAGATGCTACAGACAATCCACGAGGAGGAGGGCCAGTNGTCACC 1298  
301 LeuSerProProCysProGluPheProTyrGluIleAsnTyr 314  
|||||  
1299 CTTTCCCCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1340

seq\_name: N\_Geneseq\_36.X35013

seq\_documentation\_block:

ID X35013 standard; DNA; 1759 BP.

AC X35013;

DT 01-JUL-1999 (first entry)

DE Nucleotide sequence which has homology to PCV sequence.

KW PCV; strain PCV999; pig; PMWS; porcine multisystemic wasting syndrome;

KW vaccine; ss.

OS Sus sp.

PN PR2769321-A1.

PD 09-APR-1999.

PF 03-OCT-1997; 012382.

PR 03-OCT-1997; FR-012382.

PA (MERI-) MERIAL SAS.

PA (UYBE-) UNIV QUEENS BELFAST.

PA (UYSA-) UNIV SASKATCHEWAN.

PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,

PI Haines D, Harding J, Hassard L, Meehan B;

DR WPI; 99-246947/21.

PT New porcine circovirus from animals with porcine systemic wasting

PT syndrome

PS Disclosure: Fig 2: 35pp; French.

CC The specification describes a genomic sequence of porcine

CC circovirus (PCV) strain PCV999. The sequence is obtained from PCV

CC isolated from a physiological or tissue sample, particularly from

CC a lesion, from a pig showing symptoms of PMWS (porcine multisystemic

CC wasting syndrome), or cultured cells, infected with PCV isolated from

CC such pigs. PCV (attenuated or inactivated), polypeptides derived from

CC it, and vectors that express these polypeptides are all useful in

CC vaccines, suitable for administration to adult or young pigs, or to

CC pregnant sows (for passive immunization of their offspring). DNA

CC isolated from PCV is used for in vivo or in vitro expression of viral

CC polypeptides, also as probes or primers for diagnosis in usual

CC hybridization or amplification assays. These polypeptides may also be

CC used diagnostically to detect PCV-specific antibodies, while antibodies

CC raised against the polypeptides can be used to detect antigens, in any

CC usual immunoassay format. The present sequence appears in the

CC specification.

SQ Sequence 1759 BP; 475 A; 510 C; 342 G; 432 T;

alignment\_scores:

Quality: 1447.00 Length: 315

Ratio: 4.922 Gaps: 2

Percent Similarity: 93.333 Percent Identity: 83.810

alignment\_block:

US-09-209-961-3 x X35013/rev

Align seg 1/1 to reverse of: X35013 from: 1 to: 1759

```
1 MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTr 17
|||||
1365 ATGCCAAGCAAGAA.....AGCGGCCGCCCAACCCCAATAAGAGGTG 1325
17 pValPheThrLeuAsnAsnProSerGluAspGluArgLysLysLeArg 34
|||||
1324 GGTGTTCCACCTTAATCCTTCGAGGAGGAGAGAAACAAATACGGG 1275
34 LuLeuProIleSerLeuPheAspPheIleValGlyGluGlyAsn 50
|||||
1274 AGCTTCAACATCCTCTTTGATTATTGTTGCGGAGAGAGAGTTG 1225
51 GluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPheValLys 67
|||||
1224 GAAGAGGTAGAACTCCTCACCCTCAGGGGTTGCGAATTTGCTAGAA 1175
67 sGlnThrPheAsnLysValLysTrpTyrLeuGlyAlaArgCysHisIle 84
|||||
1174 GCAGACTTTTAAACAAGGTGAAGTGGTATTTTGGTGCCTGCCACATCG 1125
84 LuLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrCysSerLysGlu 100
|||||
1124 AGAAGCGAAAGGAACCGACGACGAGATAAAGATACTGCAGTAAGAA 1075
101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArg 117
|||||
1074 GGCACATACATTATCGAGTGTGGAGCTCCGCGGAACAGGGGAGCGCAG 1025
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuVal 134
|||||
1024 CGACCTGCTACTCTGTGAGTACCCCTTTTGGAGACGGGCTTTGGTGA 975
134 hrValAlaGluGlnHisProValThrPheValLysAsnPheArgGlyLeu 150
|||||
974 CTGTAGCGGAGCAGTCTCCCTGTAACGATGTGTGAGAAATTTCCGGGGGCTG 925
151 AlaGluLeuLeuLysValSerGlyLysMetGlnLysArgAspTrpLysTh 167
|||||
924 GCTGAACCTTTTGAAGTGAGCGGAAGATGACGAGCGGTGATGGAGAGC 875
167 rAsnValHisPheIleValGlyProProGlyCysGlyLysSerLysTrp 184
|||||
874 AGCTGTACACGTATAGTGGCGCCCGGTTGTGGAGAGCCAGTGGG 825
184 laAlaAsnPheAlaAsnProGluThrThrTyrTrpLysProLysAsn 200
|||||
824 CCCGTAATTTTGCTGACCTAGGACACCTACTGGAAGCCCTAGTAGAAT 775
201 LysTrpTrpAspGlyTyrHisGlyGluLysValValIleAspAspPh 217
|||||
774 AAGTGGTGGGATGATATCATGGAGAAGTGTGTTGTTGGATGATTT 725
217 eTyrGlyTrpLeuProTrpAspAspLeuLeuArgLysCysAspArgTyrP 234
|||||
724 TTATGGCTGTTACTTGGATGATCTACTGAGACTGTGTGACCGGTATC 675
234 roLeuThrValLysThrLysGlyGlyThrValProPheLeuAlaArgSer 250
|||||
674 CATTGACTGTAGAGACTAAAGGGGGTACTGTCTCTTTTGGCCCGCAGT 625
251 IleLeuIleThrSerAsnGlnThrProLeuGluTrpTyrSerSerThrAl 267
|||||
624 ATTTTGTATTACCAATACAGCTAGGCCCCCAGGAGTGTACTCTCAACTGC 575
267 aValProAlaValGluAlaLeuTyrArgArgIleThrSerLeuValPhe 284
|||||
574 TGTCCTCCAGCTAGAGCTCTCTATCGAGGATTAATCTTGTGCAATTTT 525
284 rPlyAsnAlaThrLysGlnSerThrGlu...GluGlyGlyGlnPheVal 299
|||||
524 GGAAGACTGTGGAGACATCAACGAGGAGTACCCGAGGCGGATTGAA 475
300 ThrLeuSerProProCysProGluPheProTyrGluIleAsnTyr 314
```

```
seq_name: N_Geneseq_36:Tl3161
seq_documentation_block:
ID Tl3161 standard; DNA; 1022 BP.
AC Tl3161;
DT 23-MAY-1996 (first entry)
DE SCSV segment 2.
KW SCSV; promoter; transcription; transgenic plant; legume;
KW gene expression; crop improvement; ss.
OS Subterranean clover stunt virus isolate F.
FH Key Location/Qualifiers
FT CDS
82..924
/*tag= a
PN WO9606932-A1.
PD 07-MAR-1996.
PF 30-AUG-1995; AU0552.
PR 07-AUG-1994; AU-007770.
PR 07-NOV-1994; AU-009281.
PA (CSIR ) COMMONWEALTH SCI & IND RES ORG.
PA (AUSU ) UNIV AUSTRALIAN NAT.
PI Boevink PJ, Chu PWG, Keese PK, Khan RI, Larkin PJ;
PI Marshall JS, Surin BP, Taylor WC, Waterhouse PM;
DR WPI; 96-160363/16
PT Circovirus transcription regulatory sequences and related constructs
PT - useful in plants, esp. leguminous plants, for the modulation of
PT gene expression
PS Claim 4; Page 58; 121pp; English.
CC The subterranean clover stunt virus (SCSV) genome has at least 7
CC distinct ssDNA components, designated segments 1-7 (Tl3160-66), each
CC contg. 1 major open reading frame and a non-coding region. Segment
CC 2 is predicted to be a viral replication-associated protein gene.
CC Genomic constructs useful in the genetic engineering of plants (esp.
CC legumes) comprise 1 or more heterologous gene(s) operatively linked
CC to a promoter region, and in some cases also a terminator region.
CC selected from segments 1-7. The transcription regulators facilitate
CC expression of foreign genes in plants and also facilitate control of
CC levels of gene expression in different plant tissue types.
SQ Sequence 1022 BP; 309 A; 183 C; 245 G; 285 T;
```

```
alignment_scores:
Quality: 176.00 Length: 319
Ratio: 1.197 Gaps: 14
Percent Similarity: 46.082 Percent Identity: 23.197

alignment_block:
US-09-209-961-3 x Tl3161 ..
Align seg 1/1 to: Tl3161 from: 1 to: 1022

15 LysArgTrpValPheThrLeuAsnAsnProSerGluAspGluArgLysLys 31
|||||
88 AGAGGTACTGTTTACATTAATTAATACGTACTGAGATAGAG..... 129
31 sileArgGluLeuProIleSerLeuPhe.....AspTyrP 43
|||||
130 ....AGAGAACATTCCTCTCCTCTTCTCTCAAGACGAATAAATATT 175
43 heileValGlyGluGlyAsnGluGlyArgThrProHisLeuGln 59
|||||
176 TCGTTGTGCGGCAGCAAACTGCAACTACTGGACAG...AAACACCTCCAG 222
60 GlyPheAlaAsnPheValLysLysGlnThrPheAsnLysValLysTrpTy 76
|||||
223 GGATTGTATCGTTCAAGAACAAATTCGTTGTTGATTGAAGAGAA 272
76 rleuGlyAlaArgCysHisIleGluLysAlaLysGlyThrAspGlnGlnA 93
|||||
273 ATTGTTATCGAGCTCACTGGGAATTCGAGAGCGCAGATTCTCAGA 322
```

```

93 snLysGluTyrCysSerLysGluClyAsnLeuLeuLeuLeuGluCysGlyAla 109
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
323 ATCGCGATTATTCGTGTAAGAA...ACCCTAATTCGTAATGGGATT 369
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
110 ProArgSerGlnGlyGlnArgSerAspLeuSerThrAlaValSerThrLe 126
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
370 CCGGTCTAGGAGGTTTGAACAAGCGGAAGC... 402
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
126 uLeuGluSerGlyLeuValThrValAlaGluGlnHisProValThr 143
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
403 .....ATGGAGATTATGAGAGGATCCCGAAGAAA 433
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
143 heValLysAsnPhaArgGlyLeuAlaGluLeuLysValSerGlyLys 159
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
434 TG.....CAATTGAAGGATCCAGATACGCTCTCGATGTGAAGCGAAG 477
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
160 MetGlnLys.....ArgAs 164
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
478 AAATTGAAGAGGAATAATTGCTCTGTATGATTTTCAGAAACTCCGTC 527
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
164 pTtpLysThrAsnValHis..... 170
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
528 ATGCGAAATTCAGCTTCACAGAGATTAATGCGGGAACACAGATCGGA 577
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
171 .....PheIleValGlyProGlyCys...GlyLysSerLysTrp 183
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
578 GTATCATCTGGTCTATGTTTCAGACGCGAGAGGAAGAAAGACGAGCTC 627
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
184 AlaAlaAsnPhaAlaAsnProGluThrThrTyrTrpLysProProLysAs 200
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
628 GCGAAGGAATTAATCAGG..... 645
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
200 nLysTrpTrpAspGlyTyrHisGlyGluLysValValValIleAspAsp 217
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
645 ..... 645
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
217 heTyrGlyTrpLeu.....ProTrpAspAspLeuLeuArg 228
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
646 ..TATGATGCTTTATACACCGGAGGAGAACCCAGGAGCTATTATAT 693
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
229 LeuCysAspArgTyrProLeuThrValLysThrLysGlyThrValPr 245
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
694 ATGTATGCTCAAGACCCA.....GAGAGGAATATCGCTTGTATGTTCC 737
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
245 oPheLeuAlaArgSerIleLeuIleThrSerAsnGlnThrProLeuGluT 262
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
738 CAGGTGTTCTCGAGATGATG.....AACTATCAGCGGATGGAGA 778
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
262 rp.....TyrSerThrAlaValProAlaValGluAla 273
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
779 TGTTGAAGAACACAGATTTTTCAGAGTACAAATATAGCCCTGTAGATCTT 828
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
274 LeuTyrArgArgIleThrSerLeuValPheTrpLysAsnAlaThrLysG 290
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
829 TGTATTAGGAAGTAGTTCATTATTAATTGTGTTTCCCAACGTTGCCACCTGA 878
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
290 nSerThr 292
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
879 CCCACG 885

```

seq\_name: N\_Geneseq\_36.V71832

seq\_documentation\_block:

```

ID V71832 standard; DNA; 1106 BP.
AC V71832;
DT 10-FEB-1999 (first entry)
DE BBTV DNA 1 clone (7-4-2) nucleotide sequence.
KW Banana bunchy top virus; BBTV; PCR technique: plant; tissue; banana;
KW Musa; phloem damage; aphid; nucleic acid detection; circular; ss.
OS Banana bunchy top virus.
FH Key Location/Qualifiers
FT TATA_signal 1..8
FT /tag= a

```

```

FT stem_loop 9..38
FT /tag= b
FT CDS 62..922
FT /tag= c
FT /product= "ORF-V2 product"
FT /transl_except= (pos:335..337, aa:Gly)
FT /transl_except= (pos:518..520, aa:Ser)
FT polyA_signal 380..385
FT /tag= d
FT polyA_signal 901..906
FT /tag= e
FT US5846705-A.
FT 08-DEC-1998.
FT 06-APR-1995; 418071.
FT 06-APR-1995; US-418071.
FT (BIOT-) DEV CENT BIOTECHNOLOGY.
FT Soong T, Wu R, You L;
FT WPI; 99-059037/05.
FT P-PSDB; W87459.
FT Nucleic acids having banana bunchy top virus component sequences -
FT used to design primers for use in polymerase chain reaction
FT detection of the virus
FT Claim 1; Fig 11A-B; 27pp; English.
CC This represents the nucleotide sequence of a banana bunchy top virus
CC (BBTV) DNA I (clone 7-4-2) circular single stranded (css) DNA. The
CC invention provides nucleic acid sequences associated with BBTV that can
CC be used in a PCR technique for detecting BBTV. The nucleic acid sequences
CC (V71830 to V71833) are used as the basis for the construction of PCR
CC primers, to detect BBTV infection. The PCR technique is used for
CC detecting BBTV in plant tissues (preferably banana, especially Musa
CC species). The virus, one of the most important banana species viruses,
CC causes phloem damage and is transmitted by aphids. PCR detection gives
CC accurate, reliable and specific determination of absence or presence of
CC the virus.
SQ Sequence 1106 BP; 335 A; 225 C; 257 G; 289 T;

```

alignment\_scores:

```

Quality: 168.50 Length: 282
Ratio: 1.170 Gaps: 13
Percent Similarity: 51.064 Percent Identity: 25.177

```

alignment\_block:

US-09-209-961-3 x V71832

Align seg 1/1 to: V71832 from: 1 to: 1106

```

13 ProHisLysArgTrpValPheThrLeuAsnAsnProSerGluAspGluar 29
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
71 CCTCTCTTAAGTGGTCTTCACTGATTAATCTCTCCGCGCAGAGAG 120
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
29 gLysLysIleArgGluLeuProIleSerLeuPhe.....A 41
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
121 AGAAACTTT.....CTCTCTCTTGAAGGAGGAGGATGTC 158
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
41 spTyrPheIleValGlyClyGluGlyAsnGluGlyArgThrProHis 57
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
159 ACTAGCTGCTCGGCGACCAAGTCTCGCGCCACCGCCGAGAACGAC 208
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
58 LeuGlnGlyPheAlaAsnPhaValLysLysGlnThrPheAsnLysVally 74
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
209 CTCGAGGATATCTATCCCTGAAAAGAGATCCGCTCGCGGATTGAA 258
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
74 sTrpTyrLeuGlyAlaArgCysHisIleGluLysAlaLysGlyThrAspG 91
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
259 GAAGAAGTATGTTCCCGTGTCTACTGGGAGATTGCCAGAGAACGACG 308
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
91 lnGlnAsnLysGluTyrCysSerLysGluGlyAsnLeuLeuIleGlyCys 107
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
309 AAGAAGATTCGAAGTACTGTTCAAAAGAA...ACCTAATTCTCGAATTA 355
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
108 GlyAlaProArgSerGlnGly...GlnArgSerAspLeuSerThrAlaVa 123
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```





```

502 CGAATGGTTTCACGAACATAAAGATGGCAAAATAAATAATCAACACA 551
168 .....AsnValHisPheIleValGlyProPro 176
552 TCGAAGTGTTCTGTATGATCGAAGATATCATCTGGGTATACGGTCCCAAC 601
177 GlyCys....GlyLysSerLysTIPalaalaAsnPheAlaAsnProGluTh 192
602 GGAGCGGAAGAAAGTCAACCTTCGCAAGATATCATCA..... 640
192 rThrTyrrTrpLysProProLysAsnLysTrpTrpAspGlyTyrrHisGlyG 209
641 .....TTAAACCC.....GGATGGGATATATCAACGGTG 671
209 LuLys.....ValValLleAspAspPheTyrrGly 219
672 GAAAGCGTCGGATATGATCATCATATACGATGATCGTCTGATATCAT 721
220 TrpLeuProTrpAspLeuLeuArgLeuCysAspArgTyrrProLeuTh 236
722 TGGATT.....ATTGATATCCCAAGAGTCATTCAGATTAT..... 757
236 rValLysThrLysGlyClyThrValProPheLeuAlaArgSerIleLeuI 253
758 .....CTGATTATGGCGTTATAGAACAAATTAAGATAGAGTTTAA 800
253 leThrSerAsnGlnThrPro 259
801 TAAATACAAATACGAACCA 820

```

seq\_name: N\_Geneseq\_36:V71834

seq\_documentation\_block:

```

ID V71834 standard; DNA; 1091 BP.
AC V71834;
DT 10-FEB-1999 (first entry)
DE BBTV DNA II clone (2) nucleotide sequence.
KW Banana bunchy top virus; BBTV; PCR technique; plant; tissue; banana;
KW Musa; phloem damage; aphid; nucleic acid detection; circular; cyclic; ss.
OS Banana bunchy top virus.
PN US5846705-A.
PD 08-DEC-1998.
PF 06-APR-1995; 418071.
PR 06-APR-1995; US-418071.
PA (BIOT-) DEV CENT BIOTECHNOLOGY.
PI Soong T, Wu R, You L;
DR WPI; 99-059037/05.
PT Nucleic acids having banana bunchy top virus component sequences -
PT used to design primers for use in polymerase chain reaction
PT detection of the virus
PS Disclosure; Fig 12A-C; 27pp; English.
CC This represents the nucleotide sequence of a banana bunchy top virus
CC (BBTV) DNA II (clone 2) and represents a subgenomic DNA. The invention
CC provides nucleic acid sequences associated with BBTV that can be used in
CC a PCR technique for detecting BBTV. The nucleic acid sequences (V71830 to
CC V71833) are used as the basis for the construction of PCR primers, to
CC detect BBTV infection. The PCR technique is used for detecting BBTV in
CC plant tissues (preferably banana, especially Musa species). The virus,
CC one of the most important banana species viruses, causes phloem damage
CC and is transmitted by aphids. PCR detection gives accurate, reliable and
CC specific determination of absence or presence of the virus.
SQ Sequence 1091 BP; 343 A; 224 C; 248 G; 276 T;

```

alignment\_scores:  
 Quality: 156.00 Length: 135  
 Ratio: 2.080 Gaps: 4  
 Percent Similarity: 55.556 Percent Identity: 34.074

alignment\_block:  
 US-09-209-961-3 x V71834 ..

Align seg 1/1 to: V71834 from: 1 to: 1091

```

16 ArgTyrValPheThrLeuAsnAsnProSerGluAspGluArgLys..... 30
83 AAATGGTCTTCACCTCTGAATATTATCTCCGCGGAGCGAGAGACTT 132
31 .....LysIleArgGluLeuProIleSerLeuPheAspTyrrP 43
133 TCTCGCTCTTCTGAAGGAGGAAGATTA.....AATTACG 167
43 helIleValGlyGluGluGlyAsnGluGluGlyArgThrProHisLeuGln 59
168 CTGTCGTCGGCGACGAAGTCTCCGAGCACCCGTCGGAACACCTCCAG 217
60 GlyPheAlaAsnPheValLysLysGlnThrPheAsnLysValLysTyrrTy 76
218 GGATATCTATCCCTGAAGAATCTATTAAAGCTTGGTGGATTGAAGAAGAG 267
76 rLeuGlyAlaArgCysHisIleGluLysAlaLysGlyThrAspGlnGlnA 93
268 GTACTCTTCGAAGGCTCCTGGGAGAGGGCGAGAGAACTGATGAACAGA 317
93 snLysGluTyrrCysSerLysGlu.....Gly 101
318 ATCCGACATATCTGTTCGAAGGAACCCCTTGAACTGGGTACTCCGGTGGT 367
102 AsnLeuLeuIleGluCysGlyAlaProArg.....SerG1 113
368 TCCTGGTTCGAAGAGCGCAAGCTTCGTAGATAGATCAGAGAGAGCCCTG 417
113 nGlyGlnArgSerAspLeuSerThrAlaValSerThrLeuLeuGluSerG 130
418 AGGAATTGAAGATGGACGATCCATCAAGTATCCAGATGCTTGGCAGTG 467
130 lyIle 131
468 GAATC 472

```

seq\_name: N\_Geneseq\_36:V24093

seq\_documentation\_block:

```

ID V24093 standard; DNA; 1105 BP.
AC V24093;
DT 11-AUG-1998 (first entry)
DE Banana bunchy top virus component 1 DNA sequence.
KW BBTV; probe; diagnostic primer; component 1; ss.
OS Banana bunchy top virus.
PN US5756708-A.
PD 26-MAY-1998.
PF 24-FEB-1994; 202186.
PR 24-FEB-1994; US-202186.
PA (UYQU-) UNIV QUEENSLAND TECHNOLOGY.
PI Burns TM, Dale JL, Harding RM, Karan M;
DR WPI; 98-321636/28.
PT Isolated Banana bunchy top virus DNA - useful as diagnostic probes
PT and primers for producing virus-resistant plants
PS Disclosure; Column 29-30; 59pp; English.
CC This sequence represents a DNA fragment from the component 1 of
CC the Banana bunchy top virus (BBTV). The DNA molecules can be used as
CC diagnostic probes or primers or can be inserted into plants or other
CC organisms e.g. to produce virus-resistant plants or to act as promoters,
CC enhancers or termination signals.
SQ Sequence 1105 BP; 359 A; 197 C; 279 G; 270 T;

```

alignment\_scores:  
 Quality: 156.00 Length: 238  
 Ratio: 1.238 Gaps: 13  
 Percent Similarity: 52.941 Percent Identity: 25.630

alignment\_block:  
 US-09-209-961-3 x V24093 ..

Align seg 1/1 to: V24093 from: 1 to: 1105

```

17 TrpValPheThrLeuAsnAsnProSerGluAspGluArgLysLysIleAar 33
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
125 TGGATGTTCCACATCACATCCGCC..... 151
33 gGluLeuProile.....SerLeuPheAspTyrPheIleValGlyGluG 48
   ::|||:::|||||:::|||||:::|||||:::|||||:::|||||:::
152 .TCACCTACCATGATGCGGAAGAGTTCAAAATATATGTTATATCAAGTGG 200
48 luGlyAsnGluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPhe 64
   || :::::|||||:::|||||:::|||||:::|||||:::|||||:::
201 AGAGGGACAGGAGGGT...ACTCGTCATGTCAGGATACGTCGAGATG 247
65 ValLysLysGlnThrPheAsnLysValLys...TrpTyrLeuGlyAlaAar 80
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::
248 AAGACAGAGAGCTCTGAGCAGATGAGAGGCTTCTCCAGCGCGCA.. 295
80 gCysHisIleGluLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrC 97
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
296 ....CACCTTGAGAAACGAAAGGAGCCAAAGAAAGACGACGCGCATACT 341
97 ysSerLysGluGlyAsnLeuLeu.....lleGluCysGlyAlaPro 110
   || |||||::: ||::: ||::: |||||::: |||||::: |||||:::
342 GTATGAAGGAAGACACAAAGATCGAAGGTCCTTCGAGTTTGGTGCCTTT 391
111 Arg.....SerGlnGlyGlnArgSe 117
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::
392 AAATGTCATGTAATGATAATTTATTTGATGTCATACAGGATATGCGTGA 441
117 rasPLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValT 134
   : :::::|||||:::|||||:::|||||:::|||||:::|||||:::
442 AACCCACAACGCGCTTGGAGTATTTATGAGTGT..... 478
134 hrValAlaGluGlnHisProValThrPheValLysAsn..... 146
   ||| |||||::: |||||::: |||||::: |||||::: |||||:::
479 .....CCAAACACCTTCGATAGAAAGTAAGATACATTA 511
147 PheArgGlyLeuAlaGluLeuLeuLysVal.....SerGlyLy 159
   ::|||:::|||||:::|||||:::|||||:::|||||:::|||||:::
512 TACAGATTCACAGCAGAGTTGTAATAAAACGAGCGGATGATAGCTGGA 561
159 sMetGlnLysArgAspTrpLysThrAsnValHisPheIleValGlyProp 176
   | :::::|||||:::|||||:::|||||:::|||||:::|||||:::
562 AACATCCTCAGTTCGGACATCGGAAGTTGAAATATTATGCGGAGC 611
176 roGlyCysGlyLysSerLysTrp.....AlaAlaAsnPheAlaAsnPro 190
   || :::::|||||:::|||||:::|||||:::|||||:::|||||:::
612 CATGTCACCGAGGATAATTTGGTCTATGCCCCAAATGGAGGAGAGGA 661
191 GluThrThrTyrTrpLys.....ProProLysAsnLysTrpTrpAs 204
   ::|||:::|||||:::|||||:::|||||:::|||||:::|||||:::
662 AAGACAACCTTATGCAAAATATTAAATGAAGACGAAAGAAATGCGTTTATTC 711
204 pGly.....TyrHisGlyGluLysV 211
   : :::::|||||:::|||||:::|||||:::|||||:::|||||:::
712 GCCAGGAGGAAATCATTTGGATATATGATGATTGTAATAATTATGAGGAA 761
211 alValValIleAsp 215
   ::|||:::|||||:::|||||:::|||||:::|||||:::|||||:::
762 TAGTATATTGAT 775

```

seq\_name: N\_Geneseq\_36.V24091

seq\_documentation\_block:

ID V24091 standard; DNA; 1103 BP.

AC V24091;

DT 11-AUG-1998 (first entry)

DE Banana bunchy top virus component 1 DNA sequence.

KW BBTV; probe; diagnostic primer; component 1; ss.

OS Banana bunchy top virus.

PN US5756708-A.

PD 26-MAY-1998.

```

PF 24-FEB-1994; 202186.
PR 24-FEB-1994; US-202186.
PA (UYQD-) UNIV QUEENSLAND TECHNOLOGY.
PI Burns TM, Dale JL, Harding RM, Karan M;
DR WPI; 98-321636/28.
PT Isolated Banana bunchy top virus DNA - useful as diagnostic probes
PT and primers and for producing virus-resistant plants
PS Disclosure; Column 25-28; 59pp; English.
CC This sequence represents a DNA fragment from the component 1 of
CC the Banana bunchy top virus (BBTV). The DNA molecules can be used as
CC diagnostic probes or primers or can be inserted into plants or other
CC organisms e.g. to produce virus-resistant plants or to act as promoters,
CC enhancers or termination signals.
SQ Sequence 1103 BP; 355 A; 192 C; 277 G; 279 T;

```

alignment\_scores:

Quality: 152.00 Length: 224  
Ratio: 1.267 Gaps: 12  
Percent Similarity: 53.571 Percent Identity: 25.000

alignment\_block:

US-09-209-961-3 x V24091 ..

Align seg 1/1 to: V24091 from: 1 to: 1103

```

17 TrpValPheThrLeuAsnAsnProSerGluAspGluArgLysLysIleAar 33
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
123 TGGATGTTCCACATCACATCCCGCT..... 149
33 gGluLeuProile.....SerLeuPheAspTyrPheIleValGlyGluG 48
   ::|||:::|||||:::|||||:::|||||:::|||||:::|||||:::
150 .TCGTACCATGATGCGGAGTGGTAAATATATGTTATATCAAGTGG 198
48 luGlyAsnGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPhe 64
   || :::::|||||:::|||||:::|||||:::|||||:::|||||:::
199 AGAGGGACAGGAGGT...ACTCGTCATGTCAGGATACGTCGAGATG 245
65 ValLysLysGlnThrPheAsnLysValLys...TrpTyrLeuGlyAlaAar 80
   ::|||:::|||||:::|||||:::|||||:::|||||:::|||||:::
246 AAGACGAAAGTTCTCTGAAACAGATGAGAGGCTTCTTCCAGCGCGCA.. 293
80 gCysHisIleGluLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrC 97
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
294 ....CACCTTGAGAAACGAAAGGAGCGCCAGGAAGACGACGCGGTACT 339
97 ysSerLysGluGlyAsnLeuLeu.....lleGluCysGlyAlaPro 110
   || |||||::: ||::: ||::: |||||::: |||||::: |||||:::
340 GTATGAAGGAAGATACAAAGATCGAAGGTCCTTCGAGTTTGGTGTCTTT 389
111 Arg.....SerGlnGlyGlnArgSe 117
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::
390 AAATGTCATGTAATGATAATTTATTTGATGTCATACAGGATATGCGTGA 439
117 rasPLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValT 134
   : :::::|||||:::|||||:::|||||:::|||||:::|||||:::
440 AACGCATAAACGCGCTCTGGAATATTTATATCATGAGTGT..... 476
134 hrValAlaGluGlnHisProValThrPheValLysAsn..... 146
   ||| |||||::: |||||::: |||||::: |||||::: |||||:::
477 .....CCGAATACCTTCGACAGAAAGTAAGATACATTA 509
147 PheArgGlyLeuAlaGluLeuLysVal.....SerGlyLy 159
   ::|||:::|||||:::|||||:::|||||:::|||||:::|||||:::
510 TACAGATGTCAGCAGAGTTTGAATAAAACGAGCGGATGATAGCTGGA 559
159 sMetGlnLysArgAspTrpLysThrAsnValHisPheIleValGlyProp 176
   | :::::|||||:::|||||:::|||||:::|||||:::|||||:::
560 GACATCCTTCAATGATGATGATCTGGAAGTGAAGTGAAGTATATGCGGAGC 609
176 ro.....GlyCysGly 179
   || :::::|||||:::|||||:::|||||:::|||||:::|||||:::
610 CATGTTATCGAAGGATTTATTTGGGTCTACGGCCCAAAATGGAGCGGAAGA 659

```



[illegible]



```
41 spTyrPheLeValGlyGluGluGlyAsnGluGluGlyArgThrProHis 57
   |||  ::|||::|||
159 ACTACGCTGTCGTCGGCGAGGTCCTCGGCCACCGCCAGAACAC 208
   |||  ::|||::|||
58 LeuGlnGlyPheAlaAsnPheValLysGlnThrPheAsnLysVally 74
   |||  ::|||::|||
209 CTCGAGGATATCTATCCCTGAAAGAGAGATCCGCTCGCGGATTGAA 258
   |||  ::|||::|||
74 stpTyrLeuGlyAlaArgCysHisIleGluLysAlaLysGlyThrAsp 91
   |||  ::|||::|||
259 GAGAGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 308
   |||  ::|||::|||
91 lncLysLysGluTyrCysSerLysGluGlyAsnLeuLeuLysGlyCys 107
   |||  ::|||::|||
309 AAGAGAAATTCGAAGTACTGTCAAAAGAA...ACCTAATTCGAATTA 355
   |||  ::|||::|||
108 GlyAlaProArgSerGlnGly...GlnArgSerAspLeuSerThrAla 123
   |||  ::|||::|||
356 GGGTTTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 405
   |||  ::|||::|||
123 lSerThrLeuLeuGluSerGlyIleLeuValThrValAlaGluGlnHis 140
   |||  ::|||::|||
406 TGCTGTCCTCTGAT.....CGCATGAAATTAAGACGC 440
   |||  ::|||::|||
140 roValThrPheValLysAsnPheArgGlyLeuAlaGluLeuLysVal 156
   |||  ::|||::|||
441 CTGAGATATT...CACAGATATCAATCTGTGAATAAGTTAAAAAATTC 487
   |||  ::|||::|||
157 SerGlyLysMet.....GlnLysArgAspTrpLys..... 166
   |||  ::|||::|||
488 AAGGAGGAGTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 537
   |||  ::|||::|||
167 .....ThrAsnValHisPheIleVal 173
   |||  ::|||::|||
538 GACGAGGCAATTCACAGAGAACCCGATGATCGAAGCATCATCTGGTCT 587
   |||  ::|||::|||
173 alGlyProProGlyCys...GlyLysSerLysTrpAlaAlaAsnPheAla 188
   |||  ::|||::|||
588 ATGCTCCTTATGTAAGGTAATCAACATATGCGAAGTCACATAATC 637
   |||  ::|||::|||
189 AsnProGluThrThrTrpLysProProLysAsnLys..... 201
   |||  ::|||::|||
638 AAGAAGGATGTTCTACACAGGCGGTGGGAAGAGAGATATCTTATT 687
   |||  ::|||::|||
202 .TrpTrpAspGlyTyrHisGlyGluLysValValIleAsp..... 215
   |||  ::|||::|||
688 CTCTATGTCGAGGAGGATCTGACAAGACATAGATATTGATATTCCTC 737
   |||  ::|||::|||
216 .....AspPheTyrGlyTrpLeuProTrpAspAspLeuLeuArg 228
   |||  ::|||::|||
738 GTTCTAATCAGGATTAT.....TTAATTATGATGTAATAGAGCA 778
   |||  ::|||::|||
229 LeuCysAspArgTyrProLeuThrValLysGlyGlyThrValPro 245
   |||  ::|||::|||
779 TTAAGGATAGGTTATAGAGAGTACTAATAACACCCATCAAGATAGT 828
   |||  ::|||::|||
245 oPheLeuAlaArg...SerIleLeuIleThrSerAsnGlnThrPro 259
   |||  ::|||::|||
829 TGAATTAGGTAATAATACATGTAATCGTCATGGCGCAATTCATGCC 874
   |||  ::|||::|||
seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq;us-08-418-071-4
```

## seq\_documentation\_block:

; Sequence 4, Application us/08418071

; Patent No. 5846705

; GENERAL INFORMATION:

; APPLICANT: Wu, Rey-Yuh

; APPLICANT: You, Li-Ru

; APPLICANT: Soong, Tai-Seng

; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA

; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS AND METHOD FOR DETECTING

; TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS

; NUMBER OF SEQUENCES: 18

```
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 5.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/418,071
; FILING DATE: 06-APR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley Jr., James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: DCB-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 596-9000
; TELEFAX: (212) 596-9090
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1096 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
; MOLECULE TYPE: genomic DNA
; DESCRIPTION: /desc "BBTV DNA II (clone 2-17)"
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Banana Bunchy Top Virus (BBTV)
; US-08-418-071-4
```

## alignment\_scores:

Quality: 163.00 Length: 290

Ratio: 1.273 Gaps: 13

Percent similarity: 44.138 Percent identity: 23.793

## alignment\_block:

US-09-209-961-3 x US-08-418-071-4 ..

Align seg 1/1 to: US-08-418-071-4 from: 1 to: 1096

```
16 ArgTrpValPheThrLeuAsnAsnProSerGluAspGluArgLysLys... 31
   ::|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
83 AAATGGTGTCTCACTCTGAATTATTCCTCCGCGGAGCGAGAGACTT 132
   ::|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
32 .....IleArgGluLeuProIleSerLeuPheAspTyrPheIleV 45
   ::|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
133 TCTCGCTCTCTCTGAAGGAGGAGGAGT.....CACTACTCTGTCTG 173
   ::|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
45 alGlyGluGluGlyAsnGluGluGlyArgThrProHisLeuGlnGlyPhe 61
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
174 TCGCGCAGGAAGTCGCTCCGCGCCACCGCCAGAGACCTCCAGGATAT 223
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
62 AlaAsnPheValLysLysGlnThrPheAsnLysValLysTrpTyrLeuG 78
   ::|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
224 CTATCCCTGAAAAAATCAATTCGCTCGCGGATTTGAAAAAGAGATATGG 273
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
78 yAlaArgCysHisIleGluLysAlaLysGlyThrAspGlnGlnAsnLysG 95
   ::|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
274 CTCCCGTCTCACTGGGAGATTGCGAAAGAGAGTGCAGACAGAAATCCCA 323
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
95 luTyrCysSerLysGluGlyAsnLeuLeuIleGluCysGlyAlaProArg 111
   ::|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
324 GATACTGTCAAGGAA...ACCCTAGTCTTGAATCGGTACTCCGGTG 370
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
112 SerGlnGlyGlnArgSerAspLeuSerThrAlaValSerThrLeuLeuG 128
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
```

alignment_scores:	
Quality:	156.00
Ratio:	2.080
Percent Similarity:	55.556
Percent Identity:	34.074
Gaps:	4
Length:	135

```
alignment_block:
US-09-209-961-3 x US-08-418-071-5 ..
Align seg 1/1 to: US-08-418-071-5 from: 1 to: 1091

16 ArgTrpValPheThrLeuAsnAsnProSerGluAspGluAurLys.....30
:::| | | | | | | | | | | | | | | | | | | | | | | | | | |
83 AAATGGTGCTTCACTCTGAATTATTCCCGCAGCGGACGAGAACACTT 132

31 .....LysIleArgGluLeuProIleSerLeuPheAspTyrP 43
| | | | | | | | | | | | | | | | | | | | | | | | | | |
133 TCTCGCCTCTTCTGAAGGAGGAAGATTA.....AAATTACG 167

43 heIeValGlYgLuGlucLYasncLuGlUglYArGTThrProHisLeuGln 59
:::| | | | | | | | | | | | | | | | | | | | | | | | | | |
168 CTGCTGTCGGCGACGAAGACTCGCTCCGAGCACCGGTCGGAAGCACTCCAG 217
```

60 GlyPheAlaAsnpheValLysLysGlnThrPheAsnLysValLysTrpTy 76

[illegible]

seq\_name: /cqn2\_6/ptodata/1/lna/5B\_COMB.seq:US-08-202-186-18

```

seq_documentation_block:
; Sequence 18, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1105 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
US-09-202-186-18

alignment_scores:
Quality: 156.00 Length: 238
Ratio: 1.238 Gaps: 13
Percent Similarity: 52.941 Percent Identity: 25.630

alignment_block:
US-09-209-961-3 x US-08-202-186-18 ..
Align seg 1/1 to: US-08-202-186-18 from: 1 to: 1105
17 TrpValPheThrLeuAsnAsnProSerGluAspGluArgLysLysIleAr 33
125 TGGATGTTCCACCAATCAATCCGCC..... 151
33 gGluLeuProIle.....SerLeuPheAspTyrPheIleValGlyGluG 48
152 .TCATCACCAGTCATCGGGAAGAGTTCAATATATGTTATATCAAGTGG 200
48 luGlyAsnGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPhe 64
201 AGAGGGGACAGAGAGGT...ACTGTCGTGTCAGGAGTACGTCGAGATG 247
65 ValLysLysGlnThrPheAsnLysValLys...TrpTyrLeuGlyAlaAr 80
248 AAGAGAGAGAGCTCTGAGCAGATGAGAGGCTCTCTCCAGGCGCA... 295
80 gCysHisIleGluLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrC 97
296 ....CACCTTGAGAAAGGAAGGAGCAAGCAAGAGCAGCGGCATACT 341
97/yysSerLysGluGlyAsnLeu.....lleGluCysGlyAlaPro 110

```

```

111 Arg.....SerGlnGlyGlnArgse 117
342 GTATGAAGGAAGACACAAAGAAATCGAAGTCCCTTCGAGTTTGGTGCCTTT 391
392 AAATTGTCATGTAATGATAATTTATTTGATGTCATACAGGATATGCGTGA 441
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValt 134
442 AACGCACAAACGCCCTTGGAGTATTATATGAGTGT..... 478
134 hrValAlaGluGlnHisProValThrPheValLysAsn..... 146
479 .....CCAAACACCTTCGATAGAACTAAGGATACATTA 511
147 PheArgGlyLeuAlaGluLeuLysVal.....SerGlyLy 159
512 TACAGAGTTCAAGCAGAGTTGAATAAACGAAGCGGATGAATAGCTGGAA 561
159 sMetGlnLysArgAspTyrLysThrAsnValHisPheIleValGlyProp 176
562 AACATCCTTCAGTTCGTGGACATCGGGAAGTTGAAATATATTATGCGGAGC 611
176 rogLyCysGlyLysSerLysTrp.....AlaAlaAsnPheAlaAsnPro 190
612 CATGTCACCGAAGGATAATTGGTCTCTATGGCCCAATGGAGGAGAAGGA 661
191 GluThrThrTyrTrpLys.....ProLysAsnLysTrpTrpAs 204
662 AAGACAACTTATCAAAATATTAAATGAAGACGAAGAAATCGTTTATTC 711
204 pGly.....TyrHisGlyGluLysV 211
712 GCCAGGAGGAAATCATGTCATATGTCATATGATATATATGAGGAAA 761
211 alValValIleAsp 215
762 TAGTTATATTGAT 775

seq_name: /cgn2_6/ptodata/1/lna/5B_COMB.seq:US-08-202-186-16

```

```

seq_documentation_block:
; Sequence 16, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399

```



```
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1103 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
US-08-202-186-16

alignment_scores:
  Quality: 152.00      Length: 224
  Ratio: 1.267        Gaps: 12
  Percent Similarity: 53.571      Percent Identity: 25.000

alignment_block:
US-09-209-961-3 x US-08-202-186-16 ..
Align seg 1/1 to: US-08-202-186-16 from: 1 to: 1103

17 TtpValPheThrLeuAsnAsnProSerGluAspGluArgLysLysIleAr 33
123 TGGATGTTCCATCAACATCCCGT..... 149
33 gGluLeuProfile.....SerLeuPheAspTyrPheIleValGlyGluG 48
150 .TCGCTACCAGTGCAGCGGATGAGTTAAATATATGTTATCAAGTGG 198
48 luGlyAsnGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPhe 64
199 AGAGGGGACAGGAGGT...ACTCGTCATGTGCAAGGATACGTCGAGATG 245
65 ValLysLysGlnThrPheAsnLysValLys...TrrTyrLeuGlyAlaAr 80
246 AAAGACGAAGTCTCTGAACAGATGAGAGGCTTCTCCAGGCGCA.. 293
80 gCysHisIleGluLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrC 97
294 ...CACCTGAGAACGAGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGG 339
97 ysSerLysGluGlyAsnLeuLeu.....IleGluCysGlyAlaPro 110
340 GTATGAAGGAGATACAAGATCGAGGTCCTTCGAGTTTGGTCTTTT 389
111 Arg.....SerGlnGlyGlnArgSe 117
390 AAATTGTCATGTAATATAATTTATTTGATGTCATACAGGATATGCGTGA 439
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValT 134
440 AACGCATAAACGGCTCTCGAATATTTATATGAGTGT..... 476
134 hrValAlaGluGlnHisProValThrPheValLysAsn..... 146
477 .....CCGATACCTTCGACAGAGTAAAGATACATTA 509
147 PheArgGlyLeuAlaGluLeuLysVal.....SerGlyLy 159
510 TACAGAGTCAAGCAGAGTTGAATAAAGAGGCGGATGAATAGCTGGAA 559
159 smetGlnLysArgSpTrpLysThrAsnValHisPheIleValGlyProp 176
560 GACATCCTTCAATGATGACATCTGAAGTAGAATAATATTATGGCGGAGC 609
176 ro.....GlyCysGly 179
610 CATGTTATCGAAGGATATTGGGTCTACGGCCCAATAGGAGCGGAAGA 659
180 LysSerLysTrpAlaAlaAsnProGluThrTyrTrpLys 196
660 AAGACAACCTTTGCAAAACATTATAGAGACTAAGATCGGTTTATTC 709
196 sPro.ProLysAsnLysTrp 202
```

```
710 GCCAGGAGGAAAAATCATTTGG 729
seq_name: /cgn2_6/ptodata/1/lna/5B_COMB.seq:US-08-202-186-17
seq_documentation_block:
; Sequence 17, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/202.186
; APPLICATION NUMBER: US/08/202.186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1104 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
US-08-202-186-17

alignment_scores:
  Quality: 152.00      Length: 224
  Ratio: 1.267        Gaps: 12
  Percent Similarity: 53.571      Percent Identity: 25.000

alignment_block:
US-09-209-961-3 x US-08-202-186-17 ..
Align seg 1/1 to: US-08-202-186-17 from: 1 to: 1104

17 TtpValPheThrLeuAsnAsnProSerGluAspGluArgLysLysIleAr 33
124 TGGATGTTCCATCAACATCCCGT..... 150
33 gGluLeuProfile.....SerLeuPheAspTyrPheIleValGlyGluG 48
151 .TCGCTACCAGTGCAGCGGATGAGTTCAATATATATGTTATCAAGTGG 199
48 luGlyAsnGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPhe 64
200 AGAGGGGACAGGAGGT...ACTCGTCATGTGCAAGGATACGTCGAGATG 246
65 ValLysLysGlnThrPheAsnLysValLys...TrrTyrLeuGlyAlaAr 80
247 AAGAGACGAAGCTCTCTGAAGCAGATGAGAGGCTTCTCCAGGCGCA.. 294
```

```
80 gCysHisIleGluLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrC 97
   |||:||||| |||:|||||:|||||: |||
295 ....CACCTTGAGAAACGAAAGGGGAGCCAGGAAGACGAGCGCTTACT 340
   |||:|||||:|||||:|||||:|||||: |||
97 ySerLysGluGlyAsnLeu.....IleGluCysGlyAlaPro 110
   |||:|||||: |||: |||: |||: |||: |||
341 GTATGAAGGAAGATACAAAGATCGAAGTCCTTCGAGTTTGGTCTTT 390
   |||: |||: |||: |||: |||: |||: |||
111 Arg.....SerGlnGlyGlnArgSe 117
   |||: |||: |||: |||: |||: |||: |||
391 AAATGTCATGTAATGATAATTTATGTCATACAGGATATCGGTGA 440
   |||: |||: |||: |||: |||: |||: |||
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValt 134
   |||: |||: |||: |||: |||: |||: |||
441 AACGCATAAAGCGCTCTGGAATATTATATGAGTGT..... 477
   |||: |||: |||: |||: |||: |||: |||
134 hrValAlaGluGlnHisProValThrPheValLysAsn..... 146
   |||: |||: |||: |||: |||: |||: |||
478 .....CCGAATACCTTCGACAGAAGTAAGGATACATTA 510
   |||: |||: |||: |||: |||: |||: |||
147 PheArgGlyLeuAlaGluLeuLysVal.....SerGlyLy 159
   |||: |||: |||: |||: |||: |||: |||
511 TACAGAGTCGACAGAGTGTGAATAAACGAGCGGATGATGCTGGA 560
   |||: |||: |||: |||: |||: |||: |||
159 sMetGlnLysArgAspTrpLysThrAsnValHisPheIleValGlyProp 176
   |||: |||: |||: |||: |||: |||: |||
561 GACATCCTCAATGCATGCAGGCTCTGAAGTAGAAAAATATTATGCGGAGC 610
   |||: |||: |||: |||: |||: |||: |||
176 ro.....GlyCysGly 179
   |||: |||: |||: |||: |||: |||: |||
611 CATGTTATCGAAGGATTATTGGTCTTCGGCCCAAAATGGAGCGGAAGGA 660
   |||: |||: |||: |||: |||: |||: |||
180 LysSerLysTrpAlaAlaAsnPheAlaAsnProGluThrTrpTrpLy 196
   |||: |||: |||: |||: |||: |||: |||
561 AAGACAGCTTTCGAACATTTATGAAGACTAAGAAATGCTTTTATTC 710
   |||: |||: |||: |||: |||: |||: |||
196 sPro.ProLysAsnLysTrp 202
   |||: |||: |||: |||: |||: |||: |||
711 GCCAGGAGGAAATCAITGG 730
```

seq\_name: /cn2\_6/ptodata/1/ina/5B\_COMB.seq:us-08-202-186-9

```
seq_documentation_block:
; Sequence 9, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
```

```
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1111 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
; US-08-202-186-9

alignment_scores:
  Quality: 147.50      Length: 305
  Ratio: 0.952        Gaps: 19
  Percent Similarity: 50.820  Percent Identity: 21.311

alignment_block:
US-09-209-961-3 x US-08-202-186-9 ..
Align seg 1/1 to: US-08-202-186-9 from: 1 to: 1111

17 TrpValPheThrLeuAsnProSer.....GluAspCl 28
   |||:|||||:|||||:|||||: |||
123 TGGATGTTCCACCATCAACATCCCACTACCACTACCACTGATGAGGATCA 172
   |||: |||: |||: |||: |||: |||: |||
28 uArgLysLysIleArgGluLeuProIleSerLeuPheAspTyrPheIleV 45
   |||: |||: |||: |||: |||: |||: |||
173 GATAAA.....TATATGGTAT 189
   |||: |||: |||: |||: |||: |||: |||
45 alGlyGluGluGlyAsnGluGluGlyArgThrProHisLeuGlnGlyPhe 61
   |||: |||: |||: |||: |||: |||: |||
190 ATCAAGTGGAGAGGGGACAGGAGGT...ACTCGTCATGTGCAAGGTTAT 236
   |||: |||: |||: |||: |||: |||: |||
62 AlaasnPheValLysLysGlnThrPheAsnLysValLys....TrpTyrLe 77
   |||: |||: |||: |||: |||: |||: |||
237 GTCGAGATGAAGACGACGAGCTCTGGAAGCAGATGAGAGGCTTCTCCC 286
   |||: |||: |||: |||: |||: |||: |||
77 uGlyAlaArgCysHisIleGluLysAlaLysGlyThrAspGlnGlnAsnL 94
   |||: |||: |||: |||: |||: |||: |||
287 AGCGCA.....CACCTTGAGAACGAAAGGAGGAGCCAGAGAGAGCGC 330
   |||: |||: |||: |||: |||: |||: |||
94 yGluTyrCysSerLysGlu..... 100
   |||: |||: |||: |||: |||: |||: |||
331 GGTCACTACTGTATGAAGAGAGATACAAAGATCGAAGGTCCTTCGAGTTT 380
   |||: |||: |||: |||: |||: |||: |||
101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArgSe 117
   |||: |||: |||: |||: |||: |||: |||
381 GGTTCATTAAATTTGTCATGT.....AATGA 406
   |||: |||: |||: |||: |||: |||: |||
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValt 134
   |||: |||: |||: |||: |||: |||: |||
407 TAATTTATTTGTCATACAGGATATCGGTGAAACG..... 443
   |||: |||: |||: |||: |||: |||: |||
134 hrValAlaGluGlnHisProValThrPheValLysAsnPhe..... 147
   |||: |||: |||: |||: |||: |||: |||
444 .....CACAAAAGCGCTTTGGAGTATTATATGATTGTCCTAACACC 485
   |||: |||: |||: |||: |||: |||: |||
148 .....ArgGlyLeuAlaGluLeuLeuLysValSerGlyLysMetGlnLy 162
   |||: |||: |||: |||: |||: |||: |||
486 TTCGATAGAAGTAAGGATACATTATACAGAGTACAAGCAGAGATGAATAA 535
   |||: |||: |||: |||: |||: |||: |||
162 s.....ArgAspTrpLysThrAsn..... 168
   |||: |||: |||: |||: |||: |||: |||
536 AACGAAGCGGATGAATAGCTGGAGAACITCTTTCAGTCGCTGGACATCAG 585
   |||: |||: |||: |||: |||: |||: |||
169 .....ValHisPheIle 172
   |||: |||: |||: |||: |||: |||: |||
586 AGGTGGAGATATCATGCGCAGCCATGTCATCGGAGATATATTTGGGTC 635
   |||: |||: |||: |||: |||: |||: |||
173 ValGlyProProGlyCys...GlyLysSerLysTrpAlaAlaAsnPheAl 188
   |||: |||: |||: |||: |||: |||: |||
636 TATGCCCAATGGAGGAGAGGAGAACAGCATGATGCAAAACATCTAAT 685
   |||: |||: |||: |||: |||: |||: |||
```

```

188 aasnProGluThrThrTyrTrpLysProProLysAsnLysTyrTrpAsp. 204
   :::::::::::::::::::: ||| ::::|
686 GAAGACGAGAAATGCGTTTATTCT...CCAGGAGAAATCATTTGGATA 732
205 .....GlyTyrHisGlyGluLysValValValIleAspAspPheTyr 218
   ||::| ||| ::::|:|:|:|:|
733 TATGTAGACTGTATAATTACGAGGATATTGTTATATTGAT..... 773
219 GlyTyrLeuPro.....TyrAspAspLeuLeuArgLeuLysCysArgTyr 233
   ::::| ||| ::::| |||
774 .....ATTCCAAGATGCAAGAGGATATTTA.....AATTA 805
233 rProLeuThrValLysGlyGlyThrVal..... 244
   ||| ::|:|:|:|:|:|:|
806 TGGGTATTAGAGGAATTAGAATGGAATATTCAAACGGGAATATG 855
245 ..ProPheLeuAla.....ArgSerIleLeuIleThrSerAsn 256
   ||| ||| ::::|:|:|:|:|
856 AACCCGTTTGAAGATAGTAGAATATGTCGAAGTCATTGTAATGGCTAAC 905
257 GlnThrProLeuGlu 261
   ||| |||
906 TTCCTTCGAGGAA 920

```

seq\_name: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:US-08-202-186-10

```

seq_documentation_block:
; Sequence 10, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1111 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
US-08-202-186-10

```

```

alignment_scores:
  Quality: 147.50      Length: 305
  Ratio: 0.952        Gaps: 19
Percent Similarity: 50.820 Percent Identity: 21.311

```

```

alignment_block:
US-09-209-961-3 x US-08-202-186-10 ..
Align seg 1/1 to: US-08-202-186-10 from: 1 to: 1111
17 TrpValPheThrLeuAsnAsnProSer.....GluAspG1 28
   ||::|:|:|:|:|:|:|:|:|:|
123 TGGATGTTCCACCATCAACAATCCACAACACTACCAGTGATGAGGGATGA 172
28 uArgLysLysIleArgGluLeuProIleSerLeuPheAspTyrPheIleV 45
   | |||
173 GATCAAA.....TACATGGTAT 189
45 aGlyGluGluGlyAsnGluGluGlyArgThrProHisLeuGlnGlyPhe 61
   |||:|:|:|:|:|:|:|:|:|
190 ATCAAGTGGAGGGGACAGAGGGT...ACTCGTCATGTGCAAGGTAT 236
62 AlaAsnPheValLysLysGlnThrPheAsnLysValLys...TrpTyrLe 77
   ::::|:|:|:|:|:|:|:|:|:|
237 GTCGAGATGAAGAGACGAAGCTCTCTGAACGACATGAGAGGCTTCTTCCC 286
77 uGlyAlaArgCysHisIleGluLysAlaLysGlyThrAspGlnGlnAsnL 94
   ||::|:|:|:|:|:|:|:|:|:|
287 AGCGCA.....CACCTTGAGAAACGAAAGGAAGCAAGCAAGAGAGCGC 330
94 ysGluTyrCysSerLysGlu..... 100
   ::::|:|:|:|:|:|:|:|:|:|
331 GGTCACTACTGTATGAAGGAGATACAAAGATCGAAGTCCCTTCGAGTTT 380
101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArgSe 117
   |||:|:|:|:|:|:|:|:|:|
381 GGTTCATTAAATGTCATGT.....AATGA 406
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValT 134
   ||::|:|:|:|:|:|:|:|:|:|
407 TAATTTATTGATGTCATACAGGATATGCGTGAACG..... 443
134 hrValAlaGluGlnHisProValThrPheValLysAsnPhe..... 147
   ::::|:|:|:|:|:|:|:|:|:|
444 .....CACAAAGGCCTTTGGAGTATTATGATTGCTCCTTAACACC 485
148 .....ArgGlyLeuAlaGluLeuLysValSerGlyLysMetGlnLys 162
   ||::|:|:|:|:|:|:|:|:|:|
486 TTCGATAGAAGTAAGGATACATTATACAGAGTACAGACAGATGAATAA 535
162 s.....ArgAspTrpLysThrAsn..... 168
   | ::::|:|:|:|:|:|:|:|:|:|
536 AACGAAGCGCATGAATAGCTGGAGAACTTCTTTTCAGTGTGGACATCAG 585
169 .....ValHisPheIle 172
586 AAGTGGAGAATATCATGGCGCAGCCCATGTCATCGGAGAATAATTGGGTC 635
173 ValGlyProGlyCys...GlyLysSerLysTrpAlaAlaAsnPheAl 188
   |||:|:|:|:|:|:|:|:|:|
636 TATGGCCCAATGGAGGAGAAAGAACACACGTATGCAAAACATCTTAAT 685
188 aAsnProGluThrThrTyrTrpLysProProLysAsnLysTyrTrpAsp. 204
   ::|:|:|:|:|:|:|:|:|:|
686 GAAGACAAGAATGCGCTTTATTCT...CCAGGAGGAATATCATTAGATA 732
205 .....GlyTyrHisGlyGluLysValValValIleAspAspPheTyr 218
   ||::|:|:|:|:|:|:|:|:|:|
733 TATGTAGACTGTATAATTACGAAGATATTGTTATATTGAT..... 773
219 GlyTyrLeuPro.....TyrAspAspLeuLeuArgLeuLysCysArgTyr 233
   ::::|:|:|:|:|:|:|:|:|:|
774 .....ATTCCAAGATGCAAGAGGATATTTA.....AATTA 805
233 rProLeuThrValLysLysGlyGlyThrVal..... 244
   ||| ::|:|:|:|:|:|:|:|:|:|
806 TGGGTTATTAGAGGAATTTAAGAGTGAATAATTCAAGCGGGAATATG 855

```

```
245 ..ProPheLeuAla.....ArgSerIleLeuIleThrSerAsn 256
||| |||
856 AACCCGTTTTCAGATAGTAGAATATGTCGAAGTCATTGTAATGCTAAC 905

257 GlnThrProLeuGlu 261
||| |||
906 TTCCTTCGGAAGGAA 920

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-202-186-24

seq_documentation_block:
; Sequence 24, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1111 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 129..989
; US-08-202-186-24

alignment_scores:
Quality: 147.50 Length: 305
Ratio: 0.952 Gaps: 19
Percent Similarity: 50.820 Percent Identity: 21.311

alignm_block:
US-09-209-961-3 x US-08-202-186-24 ..

Align seg 1/1 to: US-08-202-186-24 from: 1 to: 1111

17 TrpValPheThrLeuAsnProSer.....GluAspG1 28
|||||
150 TGGATGTTCCACATCAACATCCACACACTACCAGTCGATGAGGATGA 199

28 uArgLysLysIleArgGluLeuProIleSerLeuPheAspTyrPheIleV 45
|||
200 GATAAAA.....TATATGGTAT 216
```

```
45 alGlyGluGluYasnGluGluGlyArgThrProHisLeuGlnGlyPhe 61
||| :|||
217 ATCAAGTGGAGAGGACAGGAGGT...ACTCGTCATGTGCAAGGTTAT 263

62 AlaasnPheValLysLysGlnThrPheAsnLysValLys...TrpTyrIle 77
:|||||
264 GTCGAGATGAAGAGACGAGCTCTCTGAAGCAGATGAGAGGCTTCTCC 313

77 uGlyAlaArgCysHisIleGluLysAlaLysGlyThrAspGlnGlnAsnL 94
|||||
314 AGGCCA.....CACCTTGAGAAACGAAAGGAGCCAAAGAAAGAGCCG 357

94 ysGluTyrCysSerLysGlu..... 100
:|||||
358 GGTCACTACTGTGTAAGGAGAGATACAAGATCGAAGTCCCTTCGAGTT 407

101 GlyasnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArgSe 117
|||||
408 GGTTCATTAAATTTGTCATGT.....AATGA 433

117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValT 134
:|||||
434 TATATTATTGATGCATACAGGATATCGGTGAACG..... 470

134 hrValAlaGluGlnHisProValThrPheValLysAsnPhe..... 147
:|||||
471 .....CACAAAAGGCCTTTGGAGTATTATATGATGTCTTAACACC 512

148 .....ArgGlyLeuAlaGluLeuLysValSerGlyLysMetGlnLys 162
|||||
513 TTCGATAGAGTAAGGATACATTATACAGAGTACAAGCAGAGATGAATAA 562

162 s.....ArgAspTrpLysThrAsn..... 168
|
563 AACGAAGCGGATGAATAGCTGGAGAACTTCTTCAGTCTGTGGACATCAG 612

169 .....ValHisPheIle 172
:|||||
613 AGGTGGAGATATCATGCGCAGCCATGTCATCGGAGATATATTTGGGTC 662

173 ValGlyProProGlyCys...GlyLysSerLysTrpAlaAlaAsnPheAl 188
|||||
663 TATGCCCAATGGAGGAGAAAGAACACACGATATGCAAAACATCTAAT 712

188 aAsnProGluThrThrTyrTrpLysProLysAsnLysTrpTrpAsp. 204
:|||||
713 GAAGACGAGAAATCGTTTATTCT...CCAGGAGGAAATCATTGGATA 759

205 .....GlyTyrHisGlyGluLysValValIleAspAspPheTyr 218
|||||
760 TATGTAGACTGTATAATTACGAGGATATTGTATATTGAT..... 800

219 GlyTrpLeuPro.....TrpAspAspLeuLeuArgLeuCysAspArgTy 233
:|||||
801 .....ATTCCAAGATGCAAGAGGATTATTTA.....AATTA 832

233 rProLeuThrValLysThrLysGlyGlyThrVal..... 244
|
833 TGGGTTATTAGAGGAATTTAAGATGGAATTAATTCAAAGCGGGAATATG 882

245 ..ProPheLeuAla.....ArgSerIleLeuIleThrSerAsn 256
||| |||
883 AACCGGTTTGAAGATAGTAGAATATGTCGAAGTCATTGTAATGCTAAC 932

257 GlnThrProLeuGlu 261
||| |||
933 TTCCTTCGGAAGGAA 947

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-202-186-11
seq_documentation_block:
; Sequence 11, Application US/08202186
; Patent No. 5756708
```

```

; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1110 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
; US-08-202-186-11

```

```

alignment_scores:
  Quality: 144.50      Length: 305
  Ratio: 0.938        Gaps: 19
  Percent Similarity: 50.492  Percent Identity: 21.311

```

```

alignment_block:
US-09-209-961-3 x US-08-202-186-11 ..

```

```

Align seg 1/1 to: US-08-202-186-11 from: 1 to: 1110

```

```

17 TrpValPheThrLeuAsnAsnProSer.....GluAspG1 28
|||||:|||||:|||||:|||||:|||||:
123 TGGATGTTCCACATCAACATCCACACACTACCGATGATGAGGATGA 172
28 uArgLysLysIleArgGluLeuProIleSerLeuPheAspTyrPheIleV 45
| |||
173 GATCAAA.....TATATGGTAT 189
45 alGlyGluGluGlyAsnGluGlyArgThrProHisLeuGlnGlyPhe 61
||| :|||:||||| ||| |||:|||||:
190 ATCAAGTGGAGAGGGACAGAGGTT...ATCGTCATGTCAGAGTTAT 236
62 AlaAsnPheValLysLysGlnThrPheAsnLysValLys...TrpTyrLe 77
:|||||: :|||||: :|||||: :|||||: :|||||:
237 GTCAGATGAAGACGACGAAGCTCTCTGAAGCAGATGAGAGGCTTCTCC 286
77 uGlyAlaArgCysHisIleGluLysAlaLysGlyThrAspGlnGlnAsnL 94
||||| :|||||: |||||:|||||: |||||:|||||:
287 AGCGCA.....CACCTTGAGAACGAAGGAAGCAAGAAAGAGCGC 330
94 ysGluTyrCysSerLysGlu..... 100
:|||||: |||||
331 GGTCACTGTATGAGGAGATACAGATCGAAGGTCCTTCAGTTT 380

```

```

101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArgSe 117
||| :|||:|||||
381 GGTGATTTAAATTTGTCATGT.....AATGA 406
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValT 134
:|||||: :|||||: :|||:
407 TAAATTTATTTGATGTCATACAGGATATCGTGAACG..... 443
134 hrValAlaGluGlnHisProValThrPheValLysAsnPhe..... 147
:|||||: |||||: |||||: |||||: |||||: |||||:
444 .....CACAAAGGCGCTTTGGAGTATTATATGATTGCTCTCAACACC 485
148 .....ArgGlyLeuAlaGluLeuLysValSerGlyLysMetGlnLy 162
|||: ||| :|||:|||||: |||||: |||||: |||||: |||||:
486 TTCGATAGAAGTAAGATACATTATACAGAGTACACAGCAGAGATGAATA 535
162 s.....ArgAspTrpLysThrAsn..... 168
| :|||||: |||||: |||||: |||||: |||||: |||||: |||||:
536 AACGAAGCGATGAATAGCTGGAGAACTTCTTTCAGTCATGACATCAG 585
169 .....ValHisPheIle 172
586 AGGTGGAGAATATCATGGCGCAGCCATGTCATCGGAGAATAATTGGGTC 635
173 ValGlyProProGlyCys...GlyLysSerLysTrpAlaAlaAsnPheAl 188
||||| ||| :|||||: |||||: |||||: |||||: |||||: |||||:
636 TATGACCACAAATGGAGGAGAAAGACACGATATGCAAAACATCTAAT 685
188 aAsnProGluThrThrTyrTrpLysProProLysAsnLysTrpTrpAsp. 204
:| :|||||: ||| :|||||: ||| :|||||: |||
686 GAAGACGAGAAATGCGTTTATCT...CCAGGAGGAAATCATTTGGATA 732
205 .....GlyTyrHisGlyGluLysValValIleAspAspPheTyr 218
|||: ||| :|||||: |||||: |||||: |||||: |||||: |||||:
733 TATGTAGACTGTATAATTACGAGGATATTGTATATTGAT..... 773
219 GlyTrpLeuPro.....TrpAspAspLeuLeuArgLeuCysAspArgTy 233
|||: ||| :|||||: |||||: |||||: |||||: |||||: |||||:
774 .....ATCCAAGATGCAAGAGGATATTTA.....AATTA 805
233 rProLeuThrValLysGlyGlyThrVal..... 244
||| :|||: |||: |||||: |||||: |||||: |||||: |||||:
806 TGGGTTATTAGAGGAATTAAAGATGGAATAATCAACGCGGGAATATG 855
245 ..ProPheLeuAla.....ArgSerIleLeuIleThrSerAsn 256
||| ||| :|||||: |||||: |||||: |||||: |||||: |||||:
856 AACCGGTTTGAAGATAGTAGAATATATGCGAAGTCATTGTATGCGTAAC 905
257 GlnThrProLeuGlu 261
||| |||
906 TTCTTCGGAAGAA 920
seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-202-186-15
seq_documentation_block:
; Sequence 15, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1111 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
;
US-08-202-186-15

alignment_scores:
  Quality: 144.50      Length: 305
  Ratio: 0.932        Gaps: 20
  Percent Similarity: 50.820      Percent Identity: 21.967

alignment_block:
US-09-209-961-3 x US-08-202-186-15 ..
Align seg 1/1 to: US-08-202-186-15 from: 1 to: 1111

17 TrpValPheThrLeuAsnProSer.....GluAspG1 28
|||||:|||||:|||||:|||||:|||||:
123 TGGATGTTCCACCACTCAACATCCCAACACTACCACTGATGAGGATGA 172
|||||:|||||:|||||:|||||:|||||:
28 uArgLysLysIleArgGluLeuProIleSerLeuPheAspTyrPheIleV 45
|||||:|||||:|||||:|||||:|||||:
173 GATCAAA.....TATATGTTAT 189

45 alGlyGluGluGlyAsnGluGlyArgThrProHisLeuGlnGlyPhe 61
|||||:|||||:|||||:|||||:|||||:
190 ATCAAGTGGAGAGGGACAGAGGT...ACTGTCATGTCACAGGTTAT 236
|||||:|||||:|||||:|||||:|||||:
62 AlaAsnPheValLysLysGlnThrPheAsnLysValLys...TrpTyrLe 77
|||||:|||||:|||||:|||||:|||||:
237 GTCGAGATGAAGACCAAGCTCTCTGAAGCAGATGAGAGGCTTCTTCCC 286
|||||:|||||:|||||:|||||:|||||:
77 uGlyAlaArgCysHisIleGluLysAlaLysGlyThrAspGlnGlnAsnL 94
|||||:|||||:|||||:|||||:|||||:
287 AGCGCGCA.....CACCTTGAGAAACGAAGGGAAGCAAGAGGAGCGC 330
|||||:|||||:|||||:|||||:|||||:
94 ySGLuTyrCysSerLysGlu..... 100
|||||:|||||:|||||:|||||:|||||:
331 GGTCACTACTGATGAGAGAGATACAGATCGAGGTCCTTCGAGTTT 380
|||||:|||||:|||||:|||||:|||||:
101 GlyAsnLeuLeuLeuGluCysGlyAlaProArgSerGlnGlyGlnArgSe 117
|||||:|||||:|||||:|||||:|||||:
381 GGTGCAATTAATTGTCATCT.....AATGA 406
|||||:|||||:|||||:|||||:|||||:
117 AspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValT 134
|||||:|||||:|||||:|||||:|||||:
407 TAATTTATTTGATGTCATACAGGATATGCGTGAACG..... 443
|||||:|||||:|||||:|||||:|||||:
134 hrValAlaGluGlnHisProValThrPheVal.....LysAsn 146
|||||:|||||:|||||:|||||:|||||:
444 .....CACAAAGGCCCTTTGAGATTATATATGTTGCTCAAGACC 485
|||||:|||||:|||||:|||||:|||||:
147 Phe...ArgGlyLeuAlaGluLeuLysValSerGlyLysMetGlnLy 162
|||||:|||||:|||||:|||||:|||||:
486 TTCCATAGAGTAGGATACATATATACAGATACAGCAGAGATGAATAA 535
|||||:|||||:|||||:|||||:|||||:
;
; SEQUENCE CHARACTERISTICS:
; INFORMATION FOR SEQ ID NO: 13:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; NAME: JEFFERY, Donald D.
; ATTORNEY/AGENT INFORMATION:
; CLASSIFICATION: 435
; FILING DATE: 24-FEB-1994
; APPLICATION NUMBER: US/08/202,186
; SOFTWARE: Patent In Release #1.0, Version #1.25
; OPERATING SYSTEM: PC-DOS/MS-DOS
; COMPUTER: IBM PC compatible
; MEDIUM TYPE: Floppy disk
; COMPUTER READABLE FORM:
; ZIP: 20007-5109
; COUNTRY: USA
; CITY: Washington, D.C.
; STREET: 3000 K Street, N.W.
; ADDRESSEE: Foley & Lardner
; CORRESPONDENCE ADDRESS:
; NUMBER OF SEQUENCES: 60
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; APPLICANT: HARDING, Robert M.
; APPLICANT: DALE, James L.
; APPLICANT: BURNS, Thomas M.
; APPLICANT: KARAN, Mirko
; GENERAL INFORMATION:
; Patent No. 5756708
; Sequence 13, Application US/08202186
seq_documentation_block:
seq_name: /cgn2_6/ptodata/1/ina/5b_comb.seq:US-08-202-186-13
```

```

162 s.....ArgAspTrpLysThrAsn..... 168
|||||:|||||:|||||:|||||:|||||:
536 AACGAAGCGGATGATAGCTGGAGAACTTCTTCAGTCATGGACATCAG 585
|||||:|||||:|||||:|||||:|||||:
169 .....ValHisPheIle 172
|||||:|||||:|||||:|||||:|||||:
586 AGTGGAGATATCATGCGCAGCCATGTCATCGGAGAATAATTTGGGTC 635
|||||:|||||:|||||:|||||:|||||:
173 ValGlyProGlyCys...GlyLysSerLysTrpAlaIaAsnPheAl 188
|||||:|||||:|||||:|||||:|||||:
636 TATGCCCAATGGAGAGAAAGAACACGTATGCAAAACATCTAAT 685
|||||:|||||:|||||:|||||:|||||:
188 aAsnProGluThrTrpTyrTrpLysProLysAsnLysTrpTrpAsp. 204
|||||:|||||:|||||:|||||:|||||:
686 GAGACGAGAAATCGGTTTATCT...CCAGGAGGAAATCATTTGGATA 732
|||||:|||||:|||||:|||||:|||||:
205 .....GlyTyrHisGlyLysValValIleAspAspPheTyr 218
|||||:|||||:|||||:|||||:|||||:
733 TATGTAGACTGTATAATTACGAGGATATTGTATATTTGAT..... 773
|||||:|||||:|||||:|||||:|||||:
219 GlyTrpLeuPro.....TrpAspAspLeuLeuArgLeuCysAspArgTy 233
|||||:|||||:|||||:|||||:|||||:
774 .....ATTCCAAGATGCAAGAGGATTATTA.....AATTA 805
|||||:|||||:|||||:|||||:|||||:
233 rProLeuThrValLysThrLysGlyGlyThrVal..... 244
|||||:|||||:|||||:|||||:|||||:
806 TGGTTATTAGAGGAATTTAAGATGGAAATATTCAAAGCGGGAATATG 855
|||||:|||||:|||||:|||||:|||||:
245 ..ProPheLeuAla.....ArgSerIleLeuIleThrSerAsn 256
|||||:|||||:|||||:|||||:|||||:
856 AACCCGTTTGAAGATAGTAGAATATGCGAAGTCATTGTAATGCTAAC 905
|||||:|||||:|||||:|||||:|||||:
257 GlnThrProLeuGlu 261
|||||:|||||:|||||:|||||:|||||:
906 TTCCTTCGGAAGAA 920
|||||:|||||:|||||:|||||:|||||:
seq_name: /cgn2_6/ptodata/1/ina/5b_comb.seq:US-08-202-186-13

seq_documentation_block:
; Sequence 13, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
```

; LENGTH: 1109 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: circular  
 ; US-08-202-186-13

## alignment\_scores:

Quality: 143.50 Length: 368  
 Ratio: 0.806 Gaps: 22  
 Percent Similarity: 48.370 Percent Identity: 20.652

## alignment\_block:

US-09-209-961-3 x US-08-202-186-13 ..

Align seg 1/1 to: US-08-202-186-13 from: 1 to: 1109

```

17 TrpValPheThrLeuAsnProSer.....GluAspG1 28
|||||.....|
123 TGGATGTTTACCATCAACAATCCCAACACTACCAGTGATGAGGATGA 172
|||||.....|
28 uArgLysLysIleArgGluLeuProIleSerLeuPheAspTyrPheIleV 45
| |||
173 GATCAAA.....TACATGGTAT 189
45 alGlyGluGluGlyAsnGluGluArgThrProHisLeuGlnGlyPhe 61
|||||.....|
190 ATCAAGTGGAGGGGACAGGAGGT...ACTGTCATGTCAAGGTTAT 236
|||||.....|
62 AlaAsnPheValLysLysGlnThrPheAsnLysValLys...TrpTyrLe 77
|||||.....|
237 GTCGAGATGAAGACGAGCAAGCTCTCTGAAGCAGATGAGAGGCTTCTCC 286
|||||.....|
77 uGlyAlaArgCysHisIleLeuLysAlaLysGlyThrAspGlnGlnAsnL 94
|||||.....|
287 AGGGCCA.....CACCTTGAGAAACGAAGGAGGAGCCCAAGAAGAGCG 330
|||||.....|
94 ySGluTyrCysSerLysGlu..... 100
|||||.....|
331 GGTCACTACTGTGAAGAGGATCAAGAAATCGAAGTCCCTTCGAGTTT 380
|||||.....|
101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArgSe 117
||| ..|
381 GGTCAATTAAATGTCATG.....AATGA 406
|||||.....|
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValt 134
|||||.....|
407 TAATTTATTCATGTCATACAGGATATCGGTGAACG..... 443
|||||.....|
134 hrValAlaGluGlnHisProValThrPheValLysAsnPhe..... 147
|||||.....|
444 .....CACAAAGGCTTTGGAGTATTATATGATTGCTTAACACC 485
|||||.....|
148 .....ArgGlyLeuAlaGluLeuLysValSerGlyLysMetGlnLy 162
|||||.....|
486 TTCATAGAGTACAGGATACATTTACAGATACGACGAGATGAATAA 535
|||||.....|
162 s.....ArgAspTrpLysThrAsn..... 168
| .....|
536 AACGAAGCGGATGAATAGCTGAGAACTTCTTCAGTCCTGGACATCAG 585
|||||.....|
169 .....ValHisPheIle 172
| .....|
586 AGGTGGAGATATCATGCGCAGCATGTCATCGGAGAAATAATTGGGTC 635
|||||.....|
173 ValGlyProProGlyCys...GlyLysSerLysTrpAlaAlaAsnPheAl 188
|||||.....|
636 TATGCCCAATATGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 685
|||||.....|
188 aAsnProGluThrThrTyrTrpLysProProLysAsnLysTrpTrpAsp. 204
| .....|
686 GAACACGAGAAATCGGTTTATCT...CCAGGAGGAAATCATTTGGATA 732
|||||.....|

```

```

205 .....GlyTyrHisGlyGluLysValValValIleAspPheTyr 218
|||||.....|
733 TATGTAGACTGTATAATTACGAGGATATTGTTATATTGAT..... 773
|||||.....|
219 GlyTrpLeuPro.....TrpAspAspLeuLeuArgLeuCysAspArgTy 233
|||||.....|
774 .....ATTCCAAGATGCAAGAGGATATTATTA.....AATTA 805
|||||.....|
233 rProLeuThrValLysThrLysGlyGlyThrVal..... 244
|||||.....|
806 TGGGTATTAGAGGAATTAAGATGGATAATTCAAAGCGGGAATATG 855
|||||.....|
245 ..ProPheLeuAla.....ArgSerIleLeuIleThrSerAsn 256
|||||.....|
856 RACCCGTTTGAAGATAGTAGAATATATGTCGAAGTCATTGTAATGGCTAAC 905
|||||.....|
257 GlnThrProLeuGluTrpTyrSerSer..... 265
|||||.....|
906 TTCCTTCGAGGAAGGAATCTTTTCTGAAGATCGAATAAAGTTGGTTTC 955
|||||.....|
266 .....ThrAlaValProAlaValGluAlaLeuT 275
|||||.....|
956 TTGCTGAACAAGTAATGATCTTACAGCGCACGCTCCGACAAAGCCACT 1005
|||||.....|
275 yr.....ArgArgIleThrSerLeuValPheTrpLysAsnAlaThrLys 289
|||||.....|
1006 ATGACAAAGACAGCTGTCTGATTGACATCTGAAGCATCTAGGCGCGTA 1055
|||||.....|
280 GlnSerThrGluGluGlyGlnPheValThrLeuSerProCysPr 306
|||||.....|
1056 GGCCTGTGCAATGAACGCGCAGATCATA.....TGTC 1090
|||||.....|
306 oGlu 307
|||||
1091 CGAG 1094

seq_name: /cgn2_6/ptodata/1/ina/5b_COMB.seq:us-08-202-186-12

seq_documentation_block:
; Sequence 12, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 PIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 1111 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
US-08-202-186-12

alignment_scores:
  Quality: 143.50      Length: 305
  Ratio: 0.912        Gaps: 19
  Percent Similarity: 50.492  Percent Identity: 21.311

alignment_block:
US-09-209-961-3 x US-08-202-186-12
..
Align seg 1/1 to: US-08-202-186-12 from: 1 to: 1111

17  TrpValPheThrLeuAsnAsnProSer.....GluAspG1 28
123 TGGATGTTCCACCATCAACATCCCAACACTACCAGTGATGAGGATGA 172
28  uArgLysLysIleArgGluLeuProIleSerLeuPheAspTyrPheIleV 45
173 GATCAAA.....TATATGGTAT 189
45  alGlyGluGluGlyAsnGluGluGlyArgThrProHisLeuGlnGlyPhe 61
190 ATCAAGTGGAGAGGGACAGAGGGT...ACTGCTCATGTGCAAGGATAT 236
62  AlaAsnPheValLysLysGlnThrPheAsnLysValLys...TrpTyrLe 77
237 GTCGAGATGAAGACGAAAGCTCTCTGAAGCAGATGAGAGCTTCTTCCC 286
77  uGlyAlaArgCysHisIleGluLysAlaLysGlyThrAspGlnGlnAsnL 94
287 AGCGCGCA.....CACCTTGAGAAACGAAAGGGAAGCCAAAGNAGACGC 330
94  ysGluTyrCysSerLysGlu..... 100
331 GGTCACTACTGTATGAGGAGGATACAGAAATCGAAGTCCTTCGAGTTT 380
101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArgse 117
381 GGTGCTTTTAAATTGTCTATG.....AATGA 406
117 rasPLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValt 134
407 TAATTTATTGTGTCATACAGATATCGCTGAAACG..... 443
134 hrValAlaGluGlnHisProValThrPheValLysAsnPhe..... 147
444 .....CACAAAGCCCTCTCGAGTATTATATGTTCTCTACACC 485
148 .....ArgGlyLeuAlaGluLeuLysValSerGlyLysMetGlnLy 162
486 TTCGATAGAAGTAGGATACATTATACAGAGTACAAAGCAGAGATGATAA 535
162 s.....ArgAspTrpLysThrAsn..... 168
536 AACGAAGCGGATGAATAGTGGAGAACGCTTCTTCAGTGTGTTGGACATCAG 585
169 .....ValHisPheIle 172
586 AAGTGGAGAAATATCATGGCGCACCCATGTCATCGGAGAAATAATTGGGTC 635
173 ValGlyProGlyCys...GlyLysSerLysTrpAlaAlaAsnPheAl 188
636 TATGGCCCAATCGAGGAGAGAGAACACACCTATGCAAAACAATAAT 685
188 aAsnProGluThrThrTyrTrpLysProProLysAsnLysTrpTrpAsp. 204
686 GAAGACGAGGAATGCGTTTATTCT...CCAGGGGGAATAATCATTTGGATA 732

```

```

205 .....GlyTyrHisGlyGluLysValValValIleAspAspPheTyr 218
733 TATGTAGACTGATAATAACGAGGATATTGTATATTGAT..... 773
219 GlyTrpLeuPro.....TrpAspAspLeuLeuArgLeuCysAspArgTy 233
774 .....ATTCCAAGATGCAAGAGGATATTATTA.....AATTA 805
233 rProLeuThrValLysThrLysGlyGlyThrVal..... 244
806 TGGGTTATTAGAGAATTTAAGAATGAATTAATTCAAAGCGGGAATATG 855
245 ..ProPheLeuAla.....ArgSerIleLeuIleThrSerAsn 256
856 AACCCGTTTGAAGATAGTAGAATATGCGAAGTCATTGTATATGCTAAC 905
257 GlnThrProLeuGlu 261
906 TTCCTCCGAAGGAA 920

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-202-186-14

seq_documentation_block:
; Sequence 14, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIXE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1110 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
US-08-202-186-14

```

```

alignment_scores:
  Quality: 140.50      Length: 305
  Ratio: 0.918        Gaps: 19
  Percent Similarity: 50.164  Percent Identity: 21.311

alignment_block:
US-09-209-961-3 x US-08-202-186-14
..
Align seg 1/1 to: US-08-202-186-14 from: 1 to: 1110

```



```
17 TrpValPheThrLeuAsnProSer.....GluAspG1 28
|||||.....|
122 TGGAGTTTACCATCACAATCCACACTACCAGTATGAGGAGCA 171
|||||.....|
28 uArgLysLysIleArgGluLeuProIleSerLeuPheAspTyrPheIleV 45
|||||.....|
172 GATCAA.....TACATGGTAT 188
|||||.....|
45 alGlyGluGlyAsnGluGluGlyArgThrProHisLeuGlnGlyPhe 61
|||||.....|
189 ATCAAGTCAGAGGGGAGGAGGGT...ACTCGTCATGTGCAAGGATAC 235
|||||.....|
62 AlaAsnPheValLysLysGlnThrPheAsnLysValLys...TyrTyrIle 77
|||||.....|
236 GTGGAGATGAGAGCAGGAGCTCTCTGAGCAGATGAGAGGCTTCTTCCC 285
|||||.....|
77 uGlyAlaArgCysHisIleGluLysAlaLysGlyThrAspGlnGlnAsnL 94
|||||.....|
286 AGGCCA.....CACCTTGAGAACGAAAGGGGGCCCAAGATGAAGCC 329
|||||.....|
94 ySgluTyrCysSerLysGlu..... 100
|||||.....|
330 GGTCTACTGTATGAGGAGGATCAAGAAATCGAAGGTCCTTCGAGTT 379
|||||.....|
101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArgSe 117
|||||.....|
380 GTGCATTAATAATGTCATGT.....AATGA 405
|||||.....|
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuVal 134
|||||.....|
406 TAATTTATTTGATGTCATACAGGATATCGGTGAAG..... 442
|||||.....|
134 hrValAlaGluGlnHisProValThrPheValLysAsnPhe..... 147
|||||.....|
443 .....CACAAAGACCTTTGGAGTATTTATGATTGCTCTAATACC 484
|||||.....|
148 .....ArgGlyLeuAlaGluLeuLysValSerGlyLysMetGlnLys 162
|||||.....|
485 TTCGATGAAGTAAGGATACATTATACAGAGTACAAGCAGAAATGAATA 534
|||||.....|
162 s.....ArgAspTrpLysThrAsn..... 168
|||||.....|
535 AACGAGGGGATGATAGTGGAGAACGCTTTCAGTCTGGACATCAG 584
|||||.....|
169 .....ValHisPheIle 172
|||||.....|
585 AGGTTGAGATATCATCGCGCAGCCATGTCATCGAAGATTTATTTGGGT 634
|||||.....|
173 ValGlyProGlyCys...GlyLysSerLysTrpAlaAlaAsnPheAl 188
|||||.....|
635 TACGCCCAATGGAGGAGGAGGAAAGACACGTATGCAAAACATCTAAT 684
|||||.....|
188 asnProGluThrThrTyrTrpLysProProLysAsnLysTrpTrpAsp. 204
|||||.....|
685 GAAGACGAGATGCGTTTATCT...CCAGGAGGAAATCATTTGGATA 731
|||||.....|
205 .....GlyTyrHisGlyGluLysValValIleAspAspPheTyr 218
|||||.....|
732 TATGTAGACTGTATTAATATGAGGATATTTGTTATTTGAT..... 772
|||||.....|
219 GlyTrpLeuPro.....TrpAspAspLeuLeuArgLeuCysAspArgTy 233
|||||.....|
773 .....ATCCCTAGATGCAAGAGGATATTTA.....AATTA 804
|||||.....|
233 rProLeuThrValLysThrLysGlyGlyThrVal..... 244
|||||.....|
805 TGGTTTATAGAGGAATTAAGAATGGAATAATCAAAAGCGGGAATATG 854
|||||.....|
245 ..ProPheIleuAla.....ArgSerIleLeuIleThrSerAsn 256
|||||.....|
855 AACCCGTTTGAAGATGTAGAAATATGTCGAAGTCATTGTAATGCGTAA 904
|||||.....|
```

```
257 GlnThrProLeuGlu 261
|||||
905 TTCCTCCGAGGAA 919
|||||
seq_name: /cgn2_6/ptodata/1/lna/5A_COMB.seq:US-07-674-852-1
seq_documentation_block:
: Sequence 1, Application US/07674852
: Patent No. 5476658
: GENERAL INFORMATION:
: APPLICANT: TSAREV, SERGEI A
: APPLICANT: EMERSON, SUZANNE U
: APPLICANT: BALAYAN, MICHAEL S
: APPLICANT: PURCELL, ROBERT H
: TITLE OF INVENTION: A VACCINE AGAINST HEPATITIS A
: NUMBER OF SEQUENCES: 1
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
: STREET: Eleventh Floor, 1615 L. Street, N.W.
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20036-5601
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25;
: SOFTWARE: Word Perfect 5.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/674,852
: FILING DATE: 19910326
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: SCOTT, WATSON T
: REGISTRATION NUMBER: 26,581
: REFERENCE/DOCKET NUMBER: WTS/5683/83919/SRL
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 861-3000
: TELEFAX: (202) 822-0944
: TELEX: 6714627CUSH
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 7400 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: double
: TOPOLOGY: linear
: US-07-674-852-1

alignment_scores:
Quality: 111.00 Length: 206
Ratio: 1.156 Gaps: 12
Percent Similarity: 46.602 Percent Identity: 25.728

alignment_block:
US-09-209-961-3 x US-07-674-852-1 ..
Align seg 1/1 to: US-07-674-852-1 from: 1 to: 7400

138 GlnHisProValThrPheValLysAsnPheArgGlyLeuAlaGluLeu 154
|||||.....|
4180 CAGAAGGGAGTTGAACATCATTCAGAAATTCAGAACAGATTCATCCCTGC 4229
|||||.....|
154 ulysValSerGlyLysMetGlnLys.....ArgAspTrpL 166
|||||.....|
4230 CCAGGTCGACTCTAGTTTGTATGTCATTTGTCACCACCTGAGAGATTGTA 4279
|||||.....|
166 ysThrAsnValHis..... 170
|||||.....|
4280 TTGCTAGATCCATCAAAACITTAAGAAATTTAGCTCAATTAATCAGGCT 4329
|||||.....|
171 .....PheIleValGlyProProG1 177
|||||.....|
```

```
4330 ATGGTGACTAGGTGTAACCTGGTCTGTATTATATATGTAAGAGAGG 4379
177 yCysGlySerLysTrpAlaAlaAsnPheAla..... 188
4380 TGGAGGAAGAGTTTAACTTCTATTGTCATTGGCAACAAAAATTTGCAAAAC 4429
189 .....AsnProGlu...ThrThrTyTrpLysProProLysAsn 200
4430 ATTATGGTGTGACACAGAAAGATATATATACAAAACCTGTGCTTCA 4479
201 LysTrpTrpAspGlyTyHisGlyLysValValValIleAspAspPh 217
4480 GACTACTGGGATGATATAGTGGTCAATTGGTTGTATCATTTGATGACAT 4529
217 eTyTGlyTrp.....LeuProTrpAspAspLeuLeuArgLeuC 230
4530 T...GGTCAAAATACTACAGATGAAGATTGGTCAGATTTTGTCAATTGG 4576
230 ysAspArgTyTrpProLeu.....ThrValLysThrLysGly 241
4577 TGCTGGTTGCTCTATGAGGTAAATATGCTTCTTTGGAGAGAAAGG 4626
242 Gly.....ThrValProPheLeuAlaArgSerIleLeuIleThrSerAs 256
4627 AGACACTTTTCTTCCCGGTTT.....ATAATTGCCACATCAAA 4664
256 nGlnThrProLeuGluTrpTyr.....SerSerThrAlaValProAlav 271
4665 T.....TGGTCAAAATCCCAAGTCCCTAAGACTGTTTATGTGA 4699
271 alGluAlaLeuTyArgArgIleThrSerLeuVal..... 282
4700 AGGAGCTATAGATCGCGGCTTCATTATAGATTGAAGTCAACACGCA 4749
283 ...PheTrpLysAsnAla 287
4750 TCTTTTACAAAAATGCA 4767
```

OM of: US-09-209-961-3 to: EST:\* out\_format : pfs

Date: Dec 23, 1999 5:42 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:  
-MODEL=framet-p2n.model -DEV=xlp  
-Q=/cgn2.1/USPro.spool/US09209961/runat\_22121999\_101043\_17901/app\_query.fasta.1  
-DB=EST -OFT=fastap -SUFFIX=st -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500  
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-GAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blossum62 -TRANS=human40.cdi  
-LIST=45 -LOCALIGN=200 -THR\_SCORE=escore -ALIGN=15 -MODE=LOCAL  
-OUTWMT=pfs -NORM=stat -USER=US09209961 -NCPU=6 -ICPU=3 -NO\_XLPXY  
-WAIT -THREADS=1

Search information block:

Query: US-09-209-961-3

Query length: 314

Database: EST:\*

Database sequences: 4089388

Database length: 1713965092

Search time (sec): 1572.090000

score\_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
gb_est21:AA906356	-	92.50	175.96	716	AA906356 0194c11.s1 Soares_NFL
gb_est24:AI198072	+	86.50	171.59	346	AI198072 q154a05.x1 NCI_CGAP_B
gb_est10:AA181774	+	87.50	171.38	422	AA181774 q154d01.s1 Stratagene
gb_est1:T55734	+	87.50	170.98	436	T55734 yb40c11.s1 Stratagene fe
gb_est8:AA022948	+	87.50	168.94	516	AA022948 ze72h01.s1 Soares.feta
gb_est21:AA936814	+	85.50	167.48	405	AA936814 ch89c05.s1 NCI_CGAP_C
gb_est29:AI589903	+	85.50	166.92	295	AI589903 tm81g02.x1 NCI_CGAP_B
gb_est2:T74565	+	85.00	166.29	408	T74565 yc5d10.s1 Stratagene 11
gb_gsl1:AA532118	+	86.00	165.50	522	AA532118 ni44h07.s1 NCI_CGAP_L
gb_est1:AA226908	+	89.50	164.15	1101	AL078747 Drosophila melanogast
gb_est7:W20432	+	83.50	164.11	372	AA226908 zr19e04.s1 Stratagene
gb_gssr:AA0469737	+	83.50	163.89	379	W20432 zc8c05.s1 Soares.fetal
gb_est11:AA218011	+	85.00	163.71	505	AA0469737 CITB1-EI-2584G10.IR C
gb_est5:H95640	+	86.50	163.63	667	AI218011 mv59e06.r1 Soares.mous
gb_est24:AI221630	+	83.50	163.57	389	H95640 yv17a07.s1 Soares.fetal
gb_est11:AA243218	+	83.50	162.96	409	AI221630 gq20f10.x1 Soares.plac
gb_est15:AA525673	+	83.00	162.85	390	AA525673 EST177 Sugarcane leaf
gb_est17:AA653206	+	83.50	162.44	434	AA653206 ag64e11.s1 Gessler.Wil
gb_est9:AA070716	+	83.50	162.19	436	AA070716 z66f02.s1 Stratagene
gb_est10:AA179703	+	83.50	162.16	437	AA179703 zp53b02.s1 Stratagene
gb_est10:AA179708	+	83.50	162.08	440	AA179708 zp53b07.s1 Stratagene
gb_est11:AA205826	+	83.50	162.08	440	AA205826 zq53h03.s1 Stratagene
gb_est20:C85851	+	85.00	161.90	586	C85851 C85851 Mouse fertilized
gb_est38:AA057497	-	86.00	161.67	716	AA057497 ca04g01.x1 C elegans f
gb_est9:AA070466	+	83.50	161.41	465	AA070466 zm70b08.s1 Stratagene
gb_est5:H94863	+	84.50	161.35	560	H94863 yu57d04.s1 Soares.fetal
gb_est11:AA227144	+	83.50	161.33	468	AA227144 zr22c01.s1 Stratagene
gb_est29:AI608753	+	83.50	161.20	473	AI608753 zw90a09.x1 NCI_CGAP_H
gb_est22:AI006723	+	84.50	161.05	574	AI006723 ua67g09.r1 Soares.2NDM
gb_est31:AI688536	+	83.50	161.00	481	AI688536 wc90a06.x1 NCI_CGAP_C
gb_est9:AA070410	+	83.50	160.82	488	AA070410 qm68b12.s1 Stratagene
gb_est26:AI1371057	+	83.50	160.72	492	AI1371057 ta29g10.x1 Soares.feta
gb_est1:T556604	+	84.50	160.52	600	T556604 yb35b03.s1 Stratagene fe
gb_est23:AI135508	+	84.00	160.28	559	AI135508 GH13304.5prime GH Dros
gb_est11:AA216784	+	83.50	160.26	511	AA216784 zq96f10.s1 Stratagene
gb_est20:AA858195	+	83.50	160.12	517	AA858195 oe43b05.s1 NCI_CGAP_P
gb_est5:H79483	+	82.50	159.99	436	H79483 yu49b04.s1 Soares.fetal
gb_est1:T15728	+	83.50	159.35	551	T15728 IB1798 infant brain, Ber
gb_est31:F30063	+	80.50	159.34	320	F30063 HSPD20323 HM3 Homo sapie
gb_est18:AA070656	+	83.50	159.31	553	AA070656 ag90h01.r1 Stratagene
gb_est11:AA206124	+	83.50	159.28	554	AA206124 zq52g12.s1 Stratagene

gb\_est21:AA906004 + 82.50 159.16 3.82 467 ! AA906004 0j84b06.s1 Soares\_N  
gb\_est4:D51888 - 80.50 159.15 3.82 325 ! D51888 HM056A12B Clontech h  
gb\_est19:AA778101 + 83.50 159.13 3.83 561 ! AA778101 zf45g04.s1 Soares\_f

seq\_name: gb\_est21:AA906356

seq\_documentation\_block:  
LOCUS AA906356 716 bp mRNA EST 19-MAY-1998  
DEFINITION OJ94c11.s1 Soares\_NFL\_T\_GBC\_S1 Homo sapiens cDNA clone  
IMAGE:1505972 3' similar to gb:X05360 CELL DIVISION CONTROL PROTEIN  
2 HOMOLOG (HUMAN);, mRNA sequence.

ACCESSION AA906356

NID G3041479

VERSION AA906356.1 GI:3041479

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 716)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

JOURNAL Tumor Gene Index

COMMENT Unpublished. (1997)

On Jan 14, 1998 this sequence version replaced gi:1797807.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Trace considered overall poor quality

Insert Length: 1211 Std Error: 0.00

Seq primer: -40m13 fwd. ET from Amersham

High quality sequence stop: 1.

FEATURES

Location/Qualifiers

1..716

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone.lib="Soares\_NFL\_T\_GBC\_S1"

/lab\_host="DH10B"

/note="Organ: Pooled; Vector: pT73D-Pac (Pharmacia) with

a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;

Equal amounts of plasmid DNA from three normalized

libraries (fetal lung NBHL19W, testis NHT, and B-cell

NCI\_CGAP\_GCB1) were mixed, and ss circles were made in

vitro. Following HAP purification, this DNA was used as

tracer in a subtractive hybridization reaction. The driver

was PCR-amplified cDNAs from pools of 5,000 clones made

from the same 3 libraries. The pools consisted of

I.M.A.G.E. clones 297480-302087, 682632-687239,

726408-728711, and 729096-731399. Subtraction by Bento

Soares and M. Fatima Bonaldo.

BASE COUNT 207 a 156 c 162 g 191 t

ORIGIN

alignment\_scores:

Quality: 92.50 Length: 193

Ratio: 1.051 Gaps: 9

Percent Similarity: 45.596 Percent Identity: 20.725

alignment\_block:

US-09-209-961-3 x AA906356/rev ..

Align seg 1/1 to reverse of: AA906356 from: 1 to: 716

92 GlnAsnLysGluTyrCysSer.....LysGluGlyAs 102

688 GAATATCCGAGGAGTGCAACAAATTCGATCCCCAGTGAAGTGCATC 639

102 nLeuLeuile.....GluCysGlyAlaProArgSerGlnGlyGln.... 115



```
seq_name: gb_est10:AA181774

seq_documentation_block:
  LOCUS      AA181774      422 bp      mRNA      EST      06-JAN-1997
  DEFINITION  zp54601.s1 Stratagene NT2 neuronal precursor 937230 Homo sapiens
  CHAIN clone IMAGE:613249 3' similar to gb:J00314_rna2 TUBULIN BETA-1
  CHAIN (HUMAN);, mRNA sequence.

ACCESSION   AA181774
NID         G1765241
VERSION     AA181774.1 GI:1765241
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 422)
AUTHORS     Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chisoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
            Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
            Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
            Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,
            Trevaskis,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R.
            and Marra,M.
            Generation and analysis of 280,000 human expressed sequence tags
            Genome Res. 6 (9), 807-828 (1996)
MEDLINE     97044478
COMMENT     On May 9, 1995 this sequence version replaced gi:804184.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 312.

FEATURES             source
    source
    1..422
        /organism="Homo sapiens"
        /db_xref="GDB:4645088"
        /db_xref="taxon:9606"
        /clone="IMAGE:613249"
        /clone_lib="Stratagene NT2 neuronal precursor 937230"
        /tissue_type="neuroepithelial cells"
        /dev_stage="Ntera-2 neuroepithelial cells"
        /lab_host="SOLR (kanamycin resistant)"
        /note="Organ: brain; Vector: pBluescript SK-; Site_1:
        EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
        Oligo dt. Uninduced, exponentially growing neuroepithelial
        cells (Ntera-2/cl.D1). Average insert size: 1.0 kb;
        Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGG
        3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3'"

BASE COUNT  126 a 77 c 109 g 109 t 1 others
ORIGIN

alignment_scores:
  Quality: 87.50      Length: 100
  Ratio: 1.902      Gaps: 4
  Percent Similarity: 46.000      Percent Identity: 30.000

alignment_block:
  US-09-209-961-3 x AA181774
  Align seq 1/1 to: AA181774 from: 1 to: 422

75 TrpTyrLeuGlyAlaArgCysHisIleGluLysAlaLys..... 87
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
34 TGGCTCTTGGG...CGATGTCATCTTTTCATATGAAAGAGCAAA 80
87 ..... 87
```



```

264 ..... |||... |||
166 sThrAsnValHisPheIleValGlyProGlyCysGlyLysSerLysT 183
281 ATTAAAGAACCGTTT.....CTGGAGGCTTAGGGACCAAGGCCT 321
183 rpAlaIaAsnPhelaAsnPro..... 190
322 GGTCTCTT...TTCCTCCCTCCCAACCCCTGTATCCCTTCTCTGATT 368
190 ..... 190
369 CAGGGGAAAGGAACCTCGAGTGAAGGAAGTAGAATTTGGGAAAAGG 418
191 .....ProProlYsAsnLysT 202
419 GAANGATTCCACTTGACAGAAAGTGGGACAGACTCCTCCCAAGAGTAGAG 468
202 rPTpAspGly 205
469 CTGGAAGGGA 479

```

seq\_name: gb\_est21:AA936814

```

seq_documentation_block:
LOCUS      AA936814      405 bp      mRNA      EST      19-MAY-1998
DEFINITION      ch89c05.s1 NCI_CGAP_Co8 Homo sapiens cDNA clone IMAGE:1474184 3'
                similar to gb:J00314_rna2 TUBULIN BETA-1 CHAIN (HUMAN);, mRNA
                sequence.
ACCESSION      AA936814
NID            93094848
VERSION        AA936814.1 GI:3094848
KEYWORDS
SOURCE        human.
ORGANISM      Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
                Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 405)
AUTHORS        NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE          National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
                Tumor Gene Index
JOURNAL        Unpublished (1997)
COMMENT        On Jan 17, 1998 this sequence version replaced gi:2045665.

```

Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D.  
 DNA sequencing by: Greg Lennon, Ph.D.  
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www.bio.llnl.gov/dbp/image/image.html](http://www.bio.llnl.gov/dbp/image/image.html)

Insert Length: 990 Std Error: 0.00  
 Seq primer: -40m13 fwd. ET from Amersham  
 High quality sequence stop: 378.  
 Location/Qualifiers  
 1. .405  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="3: 9"  
 /clone="IMAGE:1474184"  
 /clone\_lib="NCI\_CGAP\_Co8"  
 /tissue\_type="adenocarcinoma"  
 /lab\_host="DH10B"  
 /note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from colon adenocarcinoma, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco

FEATURES  
 source

RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 126 a 75 c 107 g 97 t  
 ORIGIN

alignment\_scores:  
 Quality: 85.50 Length: 100  
 Ratio: 1.819 Gaps: 4  
 Percent Similarity: 47.000 Percent Identity: 29.000

alignment\_block:

US-09-209-961-3 x AA936814 ..

Align seg 1/1 to: AA936814 from: 1 to: 405

```

75 TptTyrLeuGlyAlaArgCysHisIleGluLysAlaLys..... 87
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
18 TGGCTCTTGGGG...CGATGTCATCTTTCAATATGAAAAAAGCAGCAA 64
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
87 ..... 87
65 GTTCAACACAAAATAGAAATCTCAAATGTAGGATAGACAAACCAAGTG 114
88 .....GlyThrAspGlnGlnAsnLysGluTyr.....CysSerLys 99
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
115 TGTGACGGGGGAAGCAACAGCAAAAGGAAGAAATGAGATGTTGCAAAAA 164
100 GluGlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnAr 116
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
165 GATGAGGGA.....GGTTCCCTCTCTCTGGGACTGA 199
116 GserAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuV 133
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
200 CTCAAACACTGATGTGGCAGTATACACCATTCAGAGTCAGGGGTGTCA 249
133 alThrValAlaGluGlnHisProValThrPheValLysAsnPhaArgGly 149
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
250 TTCTTTTGGGAGTAAGAAAAGTGGGATTAAATAAAACGTTCTTAGGA 299

```

seq\_name: gb\_est29:AI589903

seq\_documentation\_block:

```

LOCUS      AI589903      295 bp      mRNA      EST      21-APR-1999
DEFINITION      tm81g02.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2164562 3'
                similar to gb:J00314_rna2 TUBULIN BETA-1 CHAIN (HUMAN);, mRNA
                sequence.
ACCESSION      AI589903
NID            94598951
VERSION        AI589903.1 GI:4598951
KEYWORDS        EST.
SOURCE        human.
ORGANISM      Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
                Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 295)
AUTHORS        NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE          National Cancer Institute / National Institute of Neurological
                Disorders and Stroke, Brain Tumor Genome Anatomy Project
                (CGAP/BTGP), Tumor Gene Index
JOURNAL        Unpublished (1998)
COMMENT        On Mar 10, 1998 this sequence version replaced gi:2948781.

```

Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D., Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.







## FEATURES

Location/Qualifiers  
 1. .1101  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /clone\_lib="RPCI-96"  
 /clone="BACR4807"  
 /note="end : 77"  
 BASE COUNT 260 a 288 c 258 g 280 t 15 others  
 ORIGIN

## alignment\_scores:

Quality: 89.50 Length: 301  
 Ratio: 0.683 Gaps: 15  
 Percent Similarity: 43.522 Percent Identity: 22.259

## alignment\_block:

US-09-209-961-3 x CNS00LTZ/rev ..

Align seg 1/1 to reverse of: CNS00LTZ from: 1 to: 1101

```

1 MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTr 17
  ::::::::::::::::::::
806 CTGAACAGAGAGAGATGATGAGAGTATACCGGAGAGATCGC..... 762
17 pValPheThrLeuAsnAsnProSerGluAspGluArgLysLysIleArg 34
  ::::::::::::::::::::
761 .....TTCCCCACCGCTCGATGAAAAAGGAATGTTAGAGG 728
34 luLeuProIleSerLeuPheAspTrpPheIleValGlyGluGluGlyAsn 50
  ::::::::::::::::::::
727 AGCTG...ATCAATATGATCATCAATAGTATTATTAAAGGAACCTCCGGAA 681
51 GluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPheValLysLy 67
  ::::::::::::::::::::
680 CCACAAACCCCAAGAGACACATCTCTATAAGCCGCCCAAAAGGTCCTCCT 631
67 sGlnThrPheAsnLysValLysTrpTrpLeuGlyAlaArgCysHisIleG 84
  ::::::::::::::::::::
630 CGCG.....
84 luLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrCysSerLysGlu 100
  ::::::::::::::::::::
626 .....TCGGATCAAGAGAAT.....CTGAAGCT 603
101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArgSe 117
  ::::::::::::::::::::
602 CTGGCGCTAAGGATCAAGCGGGGCTATGAGTCGCCGAGGAGGAGGA 553
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValt 134
  ::::::::::::::::::::
552 GGATCTGAAGAGA.....GTGTAA 533
134 hrValAlaGluGlnHisProValThrPheValLysAsnPheArgGlyLeu 150
  ::::::::::::::::::::
532 CGGTGGAGGAGGAG.....GTGGCTCTCAAGACCACTTAAGCGTTA 489
151 AlaGluLeuLeuLysValSerGlyLysMetGlnLysArgAspTrpLysTh 167
  ::::::::::::::::::::
488 CCGGAC.....GGCTCCTGG..... 474
167 rAsnValHisPheIleValGly.ProProGlyCysGlyLysSerLysTrp 183
  ::::::::::::::::::::
473 .....ATCATTTGGGAACCC..... 455
184 AlaAlaAsnPheAlaAsnProGluThrThrTrpTrpLysProLysAs 200
  ::::::::::::::::::::
454 .....AACTTC.....CCCACTTGGAC 438
200 nLysTrpTrpAspGlyThrHisGlyGluLysValValIleAspAsp 217
  ::::::::::::::::::::
437 GAGGAGATGCCAAACTGGGAG.....GACTCCT 409

```

```

217 hetYrGlyTrpLeuProTrpAspAspLeuLeuArgLeuCysAspArgTyr 233
  ::::::::::::::::::::
408 GGCTGCTTGGTAGCCATGG.....CTGTTT 383
234 ProLeuThrValLysThrLysGlyGlyThrValProPheLeuAlaArgSe 250
  ::::::::::::::::::::
382 CAGCAAGATATAAGCTGTCCACCTGGA.....ATTCTGAACATAATC 342
250 rIleLeuIleThrSerAsnGlnThrProLeuGluTrpTrpTrpSerThrA 267
  ::::::::::::::::::::
341 GAGTTTACTTTGGAACTGCAACAGCTTTGGAGAAGATTACCAAACTA 292
267 laValProAlaValGluAlaLeuTyrArgArg.IleThrSerLeuValPh 283
  ::::::::::::::::::::
291 CGAGTACATATTGCTTGCCTATTGGCCAAAGAAATACCAGATATTGCCA 242
283 eTrpLysAsnAlaThrLysGlnSerThrGluGluGlyGlyGlnPheVal 299
  ::::::::::::::::::::
241 TTGGCGAGAGCAACTTCAGTTTGGAGGTGAAGAAGGTGGTCAATCCTC 193

```

seq\_name: gb\_est11:AA226908

## seq\_documentation\_block:

LOCUS AA226908 372 bp mRNA EST 11-MAR-1998  
 DEFINITION zrl19e04.s1 Stratagene NT2 neuronal precursor 937230 Homo sapiens  
 CDNA clone IMAGE:563870 3' similar to gb:J00314\_rna2 TUBULIN BETA-1  
 CHAIN (HUMAN); mRNA sequence.

## ACCESSION

AA226908

## NID

91848444

## VERSION

AA226908.1 GI:1848444

## KEYWORDS

EST.

## SOURCE

human.

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 372)

## REFERENCE

1 (bases 1 to 372)

## AUTHORS

Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisler, G., Jost, S.,

Krisman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,

Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,

Theising, B., White, Y., Wyllie, T., Waterston, R. and Wilson, R.

WashU-NCI human EST Project

Unpublished (1997)

On Sep 12, 1996 this sequence version replaced gi:1404680.

## TITLE

JOURNAL

## COMMENT

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 2293 Std Error: 0.00

Seq primer: -41m13 fwd. ET from Amersham.

## FEATURES

Location/Qualifiers

1..372

/organism="Homo sapiens"

/db\_xref="GDB:5425881"

/db\_xref="taxon:9606"

/clone="IMAGE:663870"

/tissue\_type="neuroepithelial cells"

/dev\_stage="Ntera-2 neuroepithelial cells"

/lab\_host="SOLR (kanamycin resistant)"

/notes="Organ: brain; Vector: pBluescript SK-; Site: 1;

ECORI; Site: 2: XhoI; Cloned unidirectionally. Primer:

Oligo dT. Uninduced, exponentially growing neuroepithelial

cells (Ntera-2/cl.D1). Average insert size: 1.0 kb;

Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGACGAG

3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'

BASE COUNT 109 a 74 c 91 g 98 t

ORIGIN

```

alignment_scores:
  Quality: 83.50      Length: 98
  Ratio: 1.856        Gaps: 4
  Percent Similarity: 45.918  Percent Identity: 29.592

alignment_block:
  US-09-209-961-3 x AA226908 ..

  Align seg 1/1 to: AA226908 from: 1 to: 372

75  TPTTYrLeuGlyAlaArgCysHisIleGluLysAlaLys..... 87
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
25  TGGCTCTTGGGG...CGATGTCATCTTTTCAATATGAAAAAAGCAGCAA 71
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
87  .....

72  GTTCAACACAAAATAGAAATCTCAATCTAGGATAGAACAAAACCAAGTG 121
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
88  .....GlyThrAspGlnGlnAsnLysGlu.....TyrCysSerLys 99
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
122  TGTGAGGGGGGAGCAGCAGCAAGAAAAGGAGAAATGAGATGTTGCAAAAAA 171
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
100  GluGlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnAr 116
   ::|||::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
172  GATGGAGGA.....GGGTTCCCTCTCTCTGGGGACTGA 206
   ::|||::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
116  gSerAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuV 133
   |||||::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
207  CTCAAACACTGATGTGGCAGTATACACCATTCAGAGTCAGGGGTGTTC 256
   |||||::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
133  alThrValAlaGluGlnHisProValThrPheValLysAsnPhe 147
   ::|||::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
257  TTTCTTTTGGGAGTAGAAAGGTGGGATTAGAAAGACGTTT 300
   |||||::: ||||| ||||| ||||| ||||| ||||| ||||| |||||

seq_name: gb_est7:W20432

seq_documentation_block:
  LOCUS      W20432      379 bp      mRNA      03-MAY-1996
  DEFINITION  Z668C05.s1 Soares_fetal_heart_NBHH19W Homo sapiens cDNA clone
  IMAGE:327464 3' similar to gb:J00314_rna2 TUBULIN BETA-1 CHAIN
  (HUMAN);, mRNA sequence.

ACCESSION   W20432
NID         G1295049
VERSION     W20432.1 GI:1295049
KEYWORDS    EST.
SOURCE      human.
  ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE   1 (bases 1 to 379)
AUTHORS    Hallier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevisakis,E., Waterston,R., Williamson,A., Wohlmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
  COMMENT   On Nov 29, 1993 this sequence version replaced gi:430432.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estevaton.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: mob.REGA+ET
High quality sequence stop: 325.
  FEATURES   Location/Qualifiers
            source            1..379
                                /organism="Homo sapiens"

```

```

/db_xref="GDB:1261666"
/db_xref="taxon:9606"
/clone="IMAGE:327464"
/clone_lib="Soares_fetal_heart_NBHH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: heart; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site:1: Not I; Site:2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTACCAATCTGAAGTGGAGCGCCGACATCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NBHL19W."

BASE COUNT  113 a 76 c 96 g 94 t
ORIGIN

```

```

alignment_scores:
  Quality: 83.50      Length: 98
  Ratio: 1.856        Gaps: 4
  Percent Similarity: 45.918  Percent Identity: 29.592

alignment_block:
  US-09-209-961-3 x W20432 ..

  Align seg 1/1 to: W20432 from: 1 to: 379

75  TPTTYrLeuGlyAlaArgCysHisIleGluLysAlaLys..... 87
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
18  TGGCTCTTGGGG...CGATGTCATCTTTTCAATATGAAAAAAGCAGCAA 64
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
87  .....

65  GTTCAACACAAAATAGAAATCTCAATCTAGGATAGAACAAAACCAAGTG 114
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
88  .....GlyThrAspGlnGlnAsnLysGlu.....TyrCysSerLys 99
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
115  TGTGAGGGGGGAGCAGCAGCAAGCAAGAAATGAGATGTTGCAAAAAA 164
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
100  GluGlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnAr 116
   ::|||::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
165  GATGGAGGA.....GGGTTCCCTCTCTCTGGGGACTGA 199
   ::|||::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
116  gSerAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuV 133
   |||||::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
200  CTCAAACACTGATGTGGCAGTATACACCATTCAGAGTCAGGGGTGTTC 249
   |||||::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
133  alThrValAlaGluGlnHisProValThrPheValLysAsnPhe 147
   ::|||::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
250  TTTCTTTTGGGAGTAGAAAGGTGGGATTAGAAAGACGTTT 293
   |||||::: ||||| ||||| ||||| ||||| ||||| ||||| |||||

seq_name: gb_gss7:AQ469737

```

```

seq_documentation_block:
  LOCUS      AQ469737      505 bp      DNA      22-APR-1999
  DEFINITION  CITBI-EI-2584G10.TR CITBI-EI Homo sapiens genomic clone 2584G10,
  genomic survey sequence.

ACCESSION   AQ469737
NID         Q4653391
VERSION     AQ469737.1 GI:4653391
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE   1 (bases 1 to 505)
AUTHORS    Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and

```

```

/clone_rfid="Soares mouse 3NME12.5"
/sex="unknown"
/issue_type="fetus"
/dev_stage="12.5dpc total fetus"
/lab_host="DH10B"
/notes="Organ: whole fetus; Vector: pT7M3D-Pac (
1st strand cDNA was primed with a Not I - oligo
[5' TGTTCACCAATCTGAAGTGGAGCGCGCCCTATTTCCTTTT
3'], on total mouse RNA [provided by Minoru K
State Univ.]; double-stranded cDNA was ligated
adaptors (Pharmacia), digested with Not I and c
the Not I and Eco RI sites of the modified pT7M
Library went through one round of normalization
constructed by Bento Soares and M. Fatima Bonal
BASE COUNT      151 a 168 c 208 g 140 t
ORIGIN

alignment_scores:
    Quality:      86.50      Length:    177
    Ratio:        0.951      Gaps:      8
    Percent Similarity: 51.412  Percent Identity: 26.554

alignment_block:
US-09-209-961-3 x AA218011      ..

Align seg 1/1 to: AA218011 from: 1 to: 667

66  LysLysGlnThrPheAsnLysValLysTrpTrpLeuGlyAlaArgCysHi 82
      :::::::::::::: :::: :::: ||| ::::::::::::::
124  CGGGAGGAGGCAGTGAATGCTCTGTGCTGGGCCCTGGCGCGGAGA.... 168
      :::::::::::::: :::: :::: ||| ::::::::::::::

82  silecLutLysala.LysGlnThrAspGlnGlnAsnLysGluTrpCys.... 97
      ||::: :::: ||| :::::::::::::: ||| ::::::::::::::

```

Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

```

Insert Length: 773      Std Error: 0.00
Seq primer: ml3 -40 forward.
FEATURES
    source
        1..389
            /organism="Homo sapiens"
            /db_xref="GDB:3792097"
            /db_xref="taxon:9606"
            /clone="IMAGE:242964"
            /clone_lib="Soares fetal liver spleen INFLS"
            /sex="male"
            /dev_stage="20 week-post conception fetus"
            /lab_host="DH10B (ampicillin resistant)"
            /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
            with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
            1st strand cDNA was primed with a Pac I - oligo(dT) primer
            [5', AACTGGAGAATAATTAAGATCTTTTTTTTTTTTTTTT 3'],
            double-stranded cDNA was ligated to Eco RI adaptors
            (Pharmacia), digested with Pac I and cloned into the Pac I
            and Eco RI sites of the modified pT7T3 vector. Library
            went through one round of normalization. Library
            constructed by Bento Soares and M.Fatima Bonaudo."
BASE COUNT      115 a      77 c      103 g      93 t      1 others
ORIGIN
alignment_scores:
    Quality:      83.50      Length:      98
    Ratio:        1.856      Gaps:      4
    Percent Similarity: 45.918      Percent Identity: 29.592
alignment_block:
US-09-209-961-3 x H95640 ..
Align seg 1/1 to: H95640 from: 1 to: 389
75 TrpTyrLeuGlyAlaArgCysHisIleGluLysAlaLys..... 87
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
18 TGGCTCTTGGGG..CGATGCATCTCTTTTCAATATGAAAAAAGCAGCAA 64
87 ..... 87
65 GTTCAACACAAATAGAAATCTCAATCTAGATAGACAAACCAAGTG 114
88 .....GlyThrAspGlnGlnAsnLysGlu.....TyrCysSerLys 99
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
115 TGTAGGGGGGAAGCAACAGCAAAAGGAAGAAATCAGATGTTGCAAAAAA 164
100 GluGlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnAr 116
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
165 GATGGAGGA.....GGGTTCCTCTCTCTCTGGGGACTGA 199
116 gSerAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuV 133
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
200 CTCAAACACTGATGGGAGTATACACCATCCAGAGTCAGGGGTGTCA 249
133 alThrValAlaGluGlnHisProValThrPheValLysAsnPhe 147
:: ||| :: ||| :: ||| :: ||| :: ||| :: ||| :: |||
250 TCTCTTTTGGGAGTAAGAAAGGTGGGGGATTAAAGACGCTTT 293

```



OM of: US-09-209-961-5 to: GenEmbl.\* out\_format : pfs

Date: Dec 27, 1999 2:13 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:

-MODEL=frame+2n.model -DEV=xlp  
-Q=/qgn2.1/USPTO\_spool/US09209961/runat\_22121999\_101043\_17910/app\_query.fasta.1  
-DB=GenEmbl -QMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -OGAPOP=4.500  
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blomsum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=escore -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pfs -NORM=stat -USER=US09209961 -NCPU=6 -ICPU=3 -NO\_XLPXY  
-WAIT -THREADS=1

Search information block:

Query: US-09-209-961-5

Query length: 233

Database: GenEmbl.\*

Database sequences: 780561

Database length: 2137953050

Search time (sec): 1994.760000

score\_list:

Sequence	Strd Orig	zScore	EScore	Len	Documentation
gb_vi:AF085695	- 1280.00	1907.73	2.9e-98	1768	AF085695 Porcine circovirus st
gb_vi:AF086836	- 1280.00	1907.73	2.9e-98	1768	AF086836 Porcine circovirus st
gb_vi:AF086835	- 1276.00	1901.73	6.3e-98	1768	AF086835 Porcine circovirus st
gb_vi:AF086834	- 1257.00	1873.23	2.4e-96	1768	AF086834 Porcine circovirus st
gb_vi:PCJ3185	- 1251.00	1864.23	7.7e-96	1768	AF223185 Porcine circovirus st
gb_vi:AF166528	- 1244.00	1853.73	3.0e-95	1768	AF166528 Porcine circovirus st
gb_vi:AF027217	- 1237.00	1843.23	1.1e-94	1768	AF027217 Porcine circovirus st
gb_vi:AF112862	- 1237.00	1843.23	1.1e-94	1768	AF112862 Porcine circovirus st
gb_vi:AF147751	- 1234.00	1838.73	2.0e-94	1768	AF147751 Porcine circovirus st
gb_vi:AF109397	- 1230.00	1832.73	4.4e-94	1768	AF109397 Bovine circovirus, c
gb_vi:AF154679	- 1226.00	1826.73	9.5e-94	1768	AF154679 Porcine circovirus, c
gb_vi:AF109399	- 1217.00	1813.23	5.4e-93	1768	AF109399 Porcine circovirus, c
gb_vi:AF109398	- 1207.00	1798.23	3.7e-92	1768	AF109398 Porcine circovirus ty
gb_vi:AF117753	- 1202.00	1790.73	9.6e-92	1768	AF117753 Porcine circovirus ty
gb_vi:AF118097	- 1163.00	1740.71	5.9e-89	700	AF118097 Porcine circovirus typ
gb_vi:AF012107	- 848.50	1269.51	3.3e-62	1759	AF012107 Porcine circovirus, c
gb_vi:PCOMGMN	- 834.50	1239.51	4.9e-61	1759	Y09921 Porcine circovirus, c
gb_vi:PCU49186	- 833.50	1238.01	5.9e-61	1759	U49186 Porcine circovirus, com
gb_vi:AF071879	- 820.50	1218.52	7.2e-60	1758	AF071879 Porcine circovirus, c
gb_vi:AF055392	- 692.00	1025.71	3.9e-49	1768	AF055392 Porcine circovirus ty
gb_vi:AF055391	- 681.00	1009.21	3.3e-48	1768	AF055391 Porcine circovirus Ty
gb_vi:AF055393	- 669.00	991.21	3.3e-47	1767	AF055393 Porcine circovirus Ty
gb_vi:AF055394	- 669.00	991.21	3.3e-47	1767	AF055394 Porcine circovirus Ty
gb_vi:AF135393	- 398.00	597.48	2.8e-25	438	AF135393 Porcine circovirus PCV
gb_vi:AF134203	- 365.00	548.02	1.6e-22	436	AF134203 Porcine circovirus PCV
gb_vi:AF071878	- 226.50	326.34	3.5e-10	1993	AF071878 Beak and feather dise
gb_vi:AF080560	- 214.00	307.59	3.9e-09	1993	AF080560 Beak and feather dise
gb_ba:SCU51728	+ 96.50	133.65	19.20	1549	U51728 Streptomyces coelicolor
gb_vi:BYV223589	+ 89.50	131.79	24.37	603	AJ223589 Barley yellow dwarf vi
gb_pl2:AF001525	+ 88.50	130.79	27.71	571	AF001525 Musa acuminata ripenin
gb_pl2:CN5018HA	+ 87.50	130.21	29.82	516	AL14230 Botrytis cinerea stral
gb_vi:BYDMVCPG	+ 87.00	128.08	39.20	600	X17260 Barley yellow dwarf viru
gb_vi:BYV223588	+ 86.50	127.29	43.41	603	AJ223588 Barley yellow dwarf vi
gb_vi:BYDPT	+ 85.50	125.79	52.62	603	M12347 Barley yellow dwarf viru
gb_ba:MTU13939	+ 84.50	125.40	55.30	534	U13939 Mycobacterium tuberculo
gb_pat:120874	+ 91.50	124.53	61.82	1848	I20874 Sequence 1 from patent
gb_ba:BCACSG	+ 94.00	124.40	62.88	2824	D38058 Bacillus caldotenax gen
gb_vi:BYDPCG	+ 87.50	124.30	63.66	984	X56050 Barley yellow dwarf viru
gb_vi:BYD7918	+ 84.50	124.29	63.78	603	AJ007918 Barley yellow dwarf vi
gb_vi:BYD7929	+ 84.50	124.29	63.78	603	AJ007929 Barley yellow dwarf vi
gb_pl1:ZMPCAL2	+ 90.00	123.82	67.71	1562	X57628 z.mays mRNA for polygal
gb_vi:AF141385	+ 84.00	123.72	68.58	591	AF141385 Sugarcane yellow leaf

gb\_pat:109604 + 91.50 123.70 68.79 2024 ! I09604 Sequence 4 from Pate  
gb\_pl1:ANGOX + 91.50 123.70 68.79 2024 ! X16061 Aspergillus niger go  
gb\_om:CFA388555 + 92.00 123.61 69.57 2218 ! AJ388555 Canis familiaris m

seq\_name: gb\_vi:AF085695

seq\_documentation\_block:  
LOCUS AF085695 1768 bp DNA circular VRL 30-SEP-1998  
DEFINITION Porcine circovirus strain 412, complete genome.  
ACCESSION AF085695

NID 93668362

VERSION AF085695.1 GI:3668362

KEYWORDS

SOURCE porcine circovirus.

ORGANISM porcine circovirus

REFERENCE 1 (bases 1 to 1768) Viruses; ssDNA viruses; Circoviridae; Circovirus.

AUTHORS Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.

TITLE Emergence of a new porcine circovirus

JOURNAL Unpublished

AUTHORS 2 (bases 1 to 1768)

TITLE Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.

JOURNAL Direct Submission

TITLE Submitted (21-ADG-1998) Veterinary Infectious Disease Organization,

JOURNAL University of Saskatchewan, 120 Veterinary Road, Saskatoon,

Saskatchewan S7N 5E3, Canada

FEATURES

Location/Qualifiers

1..1768

/organism="porcine circovirus"

/strain="412"

/db\_xref="taxon:46221"

51..995

/codon\_start=1

/product="P35.8"

/protein\_id="AAC61860.1"

/db\_xref="PID:g3668363"

/db\_xref="GI:3668363"

/translation="MPKNGRSGPQPKHVFLLNPNSEDERKKIREPLSLDYEFL  
VGEENGRTPHQLQGFANFVKKOTFNKVKYLGAACHIEKAKGTQDNKKYCSKEGN  
LLIECGAPRSGQSGDSLTAVSTLSEGLVTAKQHPVTVKFNRLAEALLKYSVSGKM  
LKIECKNTYFIVGPPCGSKWAANFANPETTYKPKKNKWDGYGKEKVVVDDFY  
GWLPDLDRLCLRCPLTVTKGTGTPFLARSILITNOTPLEWYSSSTAVPAVALRY  
RITSLVFWKATKSTEGGQFVLSPPCPPEFYNY"  
complement(357..671)

/codon\_start=1

/product="P12.1"

/protein\_id="AAC61865.1"

/db\_xref="PID:g3668368"

/db\_xref="GI:3668368"

/translation="MVYIPPLVFRWFPVCGFRVCKISSPFAFTTPRPHNEVYIGFPI  
TLHFPAHFQKFSQPAIEFDKRYRVLICNGHNPALQOQTHSSHQVTPLSLRSSSTF  
NK"

complement(385..565)

/codon\_start=1

/product="P6.6"

/protein\_id="AAC61866.1"

/db\_xref="PID:g3668369"

/db\_xref="GI:3668369"

/translation="MKCTLVFSQRCIFPLTFKSSAPRKELTNVWGCFATVTRIP  
SNKVLTAVDRLRCP"

553..732

/codon\_start=1

/product="P6.6"

/protein\_id="AAC61861.1"

/db\_xref="PID:g3668364"

/db\_xref="GI:3668364"

/translation="MYTSLMGLGVKANGLLLIQTRKPHGNHLKTSGGMVTWVKW  
LLMTFMAGCGRMVY"

833..970

/codon\_start=1

/product="P5.0"

/protein\_id="AAC61862.1"

/db\_xref="PID:g3668365"

/db\_xref="GI:3668365"  
/translation="NVLLNCCPSCRSLSSEDFLGLILECYKTIHGGRGVPRHPPPM  
P"

CDS

1016..1177

/codon\_start=1  
/product="P6.2"  
/protein\_id="AAC61863.1"  
/db\_xref="PID:g3668366"  
/db\_xref="GI:3668366"  
/translation="WVFIHLGFKWGVFKFSELYIHGYTDIVVLVVTVPERSAAE  
YVWHISRL"

CDS

complement(1034..1735)

/codon\_start=1  
/product="P27.9"  
/protein\_id="AAC61864.1"  
/db\_xref="PID:g3668367"  
/db\_xref="GI:3668367"  
/translation="MTYPRRRYRRRRRPSHLQILRRRPLVHPHRYRWRKNGI  
ENTRLSRFTGYTVKRTTTPSWAVDMRFKIDDFVPGGGTKLSIPPEYIRKVK  
VEFWPCSPITQDRCVGSFVILDDNFVKATALYDYPVNTSSRHTIQPFYSHSRY  
FTPKVLDSITIDFQPNRNRLWLRLQTSNVDHVLGLTAFENSKIDQDINIRVTMY  
VQREFNLKDPPLP"

BASE COUNT 463 a 362 c 481 g 462 t

ORIGIN

## alignment\_scores:

Quality: 1280.00 Length: 233  
Ratio: 5.494 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-209-961-5 x AF085695/rev ..

Align seg 1/1 to reverse of: AF085695 from: 1 to: 1768

1 MetThrProArgArgArgTyrArgArgArgArgHisArgProArgSe 17  
1735 ATGACGTATCCAGGAGCGGTACCGCAGAAAGACACCCGCCCGCAG 1686  
17 rHisLeuGlyGlnIleLeuArgArgProTrrPLeuValHisProArgH 34  
1685 CCATCTGGCAGATCTCCGCCGCCCGCTGCTGCTCACCCCGGCC 1636  
34 isArgTyrArgTrrPArgArgLysAsnGlyIlePheAsnThrArgLeuSer 50  
1635 ACCGCTACCGTTGAGAAGAAATGGCATCTTCAACACCCCGCTCC 1586  
51 ArgThrPheGlyTyrThrValLysArgThrThrValThrThrProSerTr 67  
1585 CGCACCTTCGGATATAGTCAAGCGTACCACATCAACGCCCTCGT 1536  
67 pAlaValAspMetMetArgPheLysIleAspPheValProGlyG 84  
1535 GCGGTGGACATGATGAGATTAAATGACGACTTGTTCGCCGGGAG 1486  
84 lyGlyThrAsnLysIleSerIleProPheGluTyrTyrArgIleArgLys 100  
1485 GGGGACCAACAAATCTATACCTTTGAATCTACAGATAAGAAG 1436  
101 ValLysValGluPheTrrPProCysSerProIleThrGlnGlyAspArgG 117  
1435 GTTAAGGTGAATCTCGCCTTGTCCGCCCATCACCGGGGTATAGGG 1386  
117 yValGlySerThrAlaValIleLeuAspAspAsnPheValThrLysAla 134  
1385 AGTGGGCTCCAGTCTGTTATTTAGATGATACTTTGTAAACAAGCCA 1336  
134 hrAlaLeuThrTyrAspProTyrValAsnTyrSerSerArgHisThrIle 150  
1335 CAGCCCTAACCTATGACCCATATGTAACACTACTCTCCCGCCATACAATC 1286  
151 ProGlnProPheSerTyrHisSerArgTyrPheThrProLysProValLe 167

1285 CCCAACCCCTTCTCTACCACTCCGTTACTTACACCCCAACCTGTCT 1236  
167 uAspSerThrIleAspTyrPheGlnProAsnAsnLysArgAsnGlnLeuT 184  
1235 TGATCCACTATTGATTACTTCCACCAATAAACAAGGAATCAGCTTT 1186  
184 rPLeuArgLeuGlnThrSerGlyAsnValAspHisValGlyLeuGlyThr 200  
1185 GGCTGAGGCTACAACCTCTGGAATCTGGACACGTCAGGCTCGGCAC 1136  
201 AlaPheGluAsnSerLysTyrAspGlnAspTyrAsnIleArgValThrMe 217  
1135 CGGTTCGAAACAGTAATACGACGAGCTACATATCCGTGTAACCAT 1086  
217 tTyrValGlnPheArgGluPheAsnLeuLysAspProProLeuGluPro 233  
1085 GTATGTACAATTACAGAAATTAATCTTAAAGACCCGCCCTGAACCC 1037  
seq\_name: gb\_vi:AF086836

## seq\_documentation\_block:

LOCUS AF086836 1768 bp DNA circular VRL 29-SEP-1998  
DEFINITION Porcine circovirus strain M226, complete genome.  
ACCESSION AF086836

NID 93661521

VERSION AF086836.1 GI:3661521

SOURCE porcine circovirus.

ORGANISM porcine circovirus

REFERENCE 1 (bases 1 to 1768) Viruses: ssDNA viruses; Circoviridae; Circovirus.

AUTHORS Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.

TITLE Emergence of a new porcine circovirus

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1768)

AUTHORS Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.

TITLE Direct Submission

JOURNAL Submitted (26-AUG-1998) VIDO, University of Saskatchewan, 120

Veterinary Road, Saskatoon, Saskatchewan S7N 5E3, Canada

FEATURES

source

1..1768 /organism="porcine circovirus"

/strain="M226"

/db\_xref="taxon:46221"

51..995

/codon\_start=1

/product="P35.8"

/protein\_id="AAC61741.1"

/db\_xref="PID:g3661522"

/db\_xref="GI:3661522"

/translation="MPSKNGRSGQPQHRKRWFTLNNPSEDERKKIRELPISLDFYFI  
VGEENEGRTPHLOGFANFVKOTENKVKYFGARCHIEKAKGTQONKEYSKEGN  
LLIECGAPRSQGSRLSTAVSTLLESGILTVAKOHPVTFVKNFRGLAELLKVSCKM  
QKRDWNTVHFIVGPPCGSKSWAANFANPETTYWKPPKNKWDGYHGEKVVYIDDFY  
GMLPDDLLRLCDRIPLTVTKIGTVFFLARSILITSNQTPLEWISSTAVPAVEALYR  
RITSLVFWKNATKQSTEGGQFVTLSPPCPEFFEYNY"  
complement(1034..1735)

CDS

/codon\_start=1

/product="P27.9"

/protein\_id="AAC61742.1"

/db\_xref="PID:g3661523"

/db\_xref="GI:3661523"

/translation="MTYPRRRYRRRRRPSHLQILRRRPLVHPHRYRWRKNGI  
ENTRLSRFTGYTVKRTTTPSWAVDMRFKIDDFVPGGGTKLSIPPEYIRKVK  
VEFWPCSPITQDRCVGSFVILDDNFVKATALYDYPVNTSSRHTIQPFYSHSRY  
FTPKVLDSITIDFQPNRNRLWLRLQTSNVDHVLGLTAFENSKIDQDINIRVTMY  
VQREFNLKDPPLP"

BASE COUNT 463 a 363 c 480 g 462 t

ORIGIN

alignment\_scores:



Quality: 1280.00 Length: 233  
Ratio: 5.494 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-209-961-5 x AF086836/rev ..

Align seg 1/1 to reverse of: AF086836 from: 1 to: 1768

1 MetThrTyrProArgArgArgTyrArgArgArgHisArgProArgSe 17  
|||||  
1735 ATGACGTATCCAGGAGGCGTTACCCGAGAGAGACACGCCGCCGCGAG 1686  
17 rHisLeuGlyGlnIleLeuArgArgArgProTrrLeuValHisProArgH 34  
|||||  
1685 CCATCTTGGCCAGATCTCGCGCGCGCCCTGGCTGCTCACCCGCCCTCC 1636  
34 IsArgTyrArgTrrArgArgLysAsnGlyIlePheAsnThrArgLeuSer 50  
|||||  
1635 ACCGCTACCGTTGGAGAGGAAAAATGGCATCTTCAACACCCGCCCTCC 1586  
51 ArgThrPheGlyTyrThrValLysArgThrThrValThrThrProSerTr 67  
|||||  
1585 CGCACCCTTCGGATATCTGTCAGCGTACACAGTACACAGCGCCCTCTG 1536  
67 pAlaValAspMetMetArgPheLysIleAspAspPheValProGlyG 84  
|||||  
1535 GCGGTGGACATGATGAGATTAAATGACGACTTTGTTCCCGCGGAG 1486  
84 IyGlyThrAsnLysIleSerIleProPheGlyTyrTyrArgIleArgLys 100  
|||||  
1485 GGGGACCAACAAATCTCTATACCTTTGAATACTACAGAAAGAAAG 1436  
101 ValLysValGluPheTrrProCysSerProIleThrGlnGlyAspArgG 117  
|||||  
1435 GTTAAGGTGAATCTTCGCGCTTGTCTCCCATCACCGAGGTGATAGGG 1386  
117 yValGlySerThrAlaValIleLeuAspAspAsnPheValThrLysala 134  
|||||  
1385 AGTGGGCTCCATCGCTGTTATTTAGATGATAACTTTGTAACAAAGCCA 1336  
134 hrAlaLeuThrTyrAspProTrrValAsnTyrSerSerArgHisThrIle 150  
|||||  
1335 CAGCCCTAACCTATGACCATATGTAATGTAATCTCTCCCGCCATACATC 1286  
151 ProGlnProPheSerTrrHisSerArgTyrPheThrProLysProValle 167  
|||||  
1285 CCCAACCCCTTCTCTACCACTCCCGTTACTTCCACACCCAAACCTGTCT 1236  
167 uAspSerThrIleAspTyrPheGlnProAsnAsnLysArgAsnGlnLeu 184  
|||||  
1235 TGACTCCACTATTGATTACTTCCACCAAAATACAAAGGAATCAGCTTT 1186  
184 tPLeuArgLeuGlnThrSerGlyAsnValAspHisValGlyLeuGlyThr 200  
|||||  
1185 GCGTGGGCTACAAACCTCTGAAATGTGGACCATGAGCCCTCGGCAT 1136  
201 AlaPheGluAsnSerLysTyrAspGlnAspTyrAsnIleArgValThrMe 217  
|||||  
1135 GCGTTCGAAACAGTAATACGACGAGGACTCAATATCGGTGAACAT 1086  
217 tTyrValGlnPheArgGluPheAsnLeuLysAspProLeuGluPro 233  
|||||  
1085 GPATGTACAATTCAGAGATTAAATCTTAAAGACCCCCCTGTAACCC 1037

seq\_name: gb\_v1:AF086835

seq\_documentation\_block:

LOCUS AF086835 1768 bp DNA circular VRL 29-SEP-1998  
DEFINITION Porcine circovirus strain 9741, complete genome.  
ACCESSION AF086835  
NID 93661518  
VERSION AF086835.1 GI:3661518

## KEYWORDS

SOURCE porcine circovirus.

ORGANISM

Viruses; ssDNA viruses; Circoviridae; Circovirus.

REFERENCE 1 (bases 1 to 1768)

AUTHORS Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.

TITLE Emergence of a new porcine circovirus

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1768)

AUTHORS Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.

TITLE Direct Submission

JOURNAL Submitted (26-AUG-1998) VIDO, University of Saskatchewan, 120

Veterinary Road, Saskatoon, Saskatchewan S7N 5E3, Canada

## FEATURES

source

1..1768

/organism="porcine circovirus"

/strain="9741"

/db\_xref="taxon:46221"

51..995

/codon\_start=1

/product="P35.8"

/protein\_id="AAC61739.1"

/db\_xref="PID:G3661519"

/db\_xref="GI:3661519"

/translation="MPSKKNRSGPQPHKRWVTLNPNSENKRKKIRELPISLFDYFI  
VCEGNEGRTPLHGFANFVKOTENKVKYLGARCHIEKAKGTDQONKEYCSKEGN  
LLIEGAPRSQCQSDLSSTVLESGLIVAKQHPVTFVKNRGLAELLKVSCKM  
QRDWNKTVHFIVGPPGCKSKWAANFNPETTYWPKPKNWDGYHGEKVVVDDFY  
GWLWDLLRLCDRYPLTVKTKGTVPFLARSLITSNOTPLEWYSSTAVPAVEALYR  
RITSVFWKNATEQSTEGQFVTLSPPCPEPEYIN"

## CDS

complement(1034..1735)

/codon\_start=1

/product="P27.9"

/protein\_id="AAC61740.1"

/db\_xref="PID:G3661520"

/db\_xref="GI:3661520"

/translation="NTYPRRYRRHRHRSRLGOLLRRPWLHPHRYRWRKNGI  
FNTRLSPFGTYVKRTTWTTPSWAVDMRFKIDDFVPGCGTNNKISIPFYRIRKY  
VEFWPCSPITGDRGVSTAVLDDNFVTRKATLYDPVYNNYSRHTIPQPPSYHSRY  
FTPKPVLDSIDYFQPNKRNQLRLQTSNVDHVLGTAFENSKYDQDYNIRVTMY  
VQREFNLKDPPLKP"

## BASE COUNT

464 a 360 c 480 g 464 t

## ORIGIN

## alignment\_scores:

Quality: 1276.00 Length: 233

Ratio: 5.476 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 99.571

## alignment\_block:

US-09-209-961-5 x AF086835/rev ..

Align seg 1/1 to reverse of: AF086835 from: 1 to: 1768

1 MetThrTyrProArgArgArgTyrArgArgArgHisArgProArgSe 17  
|||||  
1735 ATGACGTATCCAGGAGGCGTTACCCGAGAGAGACACGCCGCCGCGAG 1686  
17 rHisLeuGlyGlnIleLeuArgArgArgProTrrLeuValHisProArgH 34  
|||||  
1685 CCATCTTGGCCAGATCTCGCGCGCGCCCTGGCTGCTCACCCGCCCTCC 1636  
34 IsArgTyrArgTrrArgArgLysAsnGlyIlePheAsnThrArgLeuSer 50  
|||||  
1635 ACCGCTACCGTTGGAGAGGAAAAATGGCATCTTCAACACCCGCCCTCC 1586  
51 ArgThrPheGlyTyrThrValLysArgThrThrValThrThrProSerTr 67  
|||||  
1585 CGCACCCTTCGGATATCTGTCAGCGTACACAGTACACAGCGCCCTCTG 1536  
67 pAlaValAspMetMetArgPheLysIleAspAspPheValProGlyG 84  
|||||

```

1535 GCGGTGACATGATGAGATTAAATTCACGACTTGTTCCTCCCGGAG 1486
84 1yGlyThrAsnLysIleSerIleProPheGluTyrTyrArgIleArgLys 100
1485 GGGGACCAACAATAATCTATACCTTTGAATACACAGATAAGAAAG 1436
101 ValLysValGluPheTyrProCysSerProIleThrGlnGlyAspArgG1 117
1435 GTTAAGTTGAATCTGCGCTGCTCCCATCACCCAGGGGTAGGGG 1386
117 yValGlySerThrAlaValIleLeuAspAspAsnPheValThrLysAlat 134
1385 AGTGGCTCCACTGCTGTTATTTAGATGATAACTTTGTAACAAAGGCCA 1336
134 hrAlaLeuThrTyrAspProTyrValAsnTyrSerSerArgHisThrIle 150
1335 CAGCCCTAACCTATGACCCATATGTAACACTACTCTCCCGCCATACAATC 1286
151 ProGlnProPheSerTyrHisSerArgTyrPheThrProLysProValle 167
1285 CCCCACCTTCTCTACCACTCCCGTTACTTACACCCAAACCTGTCT 1236
167 uAspSerThrIleAspTyrPheGlnProAsnAsnLysArgAsnGlnLeu 184
1235 TGACTCCACTATTGATTACTTCCAAACCAATAAACAAGGAATCAGCTTT 1186
184 rPLeuArgLeuGlnThrSerGlyAsnValAspHisValGlyLeuGlyThr 200
1185 GGCTGAGCTACAAACCTCTGGAATGTGACACAGCTAGGCTCGGCAC 1136
201 AlaPheGluAsnSerLysTyrAspGlnAspTyrAsnIleArgValThrMe 217
1135 CGTTCGAAACAGTAATACAGCAGGACTACATATCCGTTGTAACCAT 1086
217 tTyrValGlnPheArgGluPheAsnLeuLysAspProProLeuGluPro 233
1085 GTATGTACAATTCAGAGAATTAACTTAAAGACCCCCCACTTAACCC 1037
seq_name: gb_v1:AF086834

seq_documentation_block:
LOCUS AF086834 1768 bp DNA circular VRL 29-SEP-1998
DEFINITION Porcine circovirus strain B9, complete genome.
ACCESSION AF086834
NID 33661515
VERSION AF086834.1 GI:3661515
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 1768)
AUTHORS Wang,L., Willson,P., Chow,B., Gibbons,E. and Babiuk,L.
TITLE Emergence of a new porcine circovirus
JOURNAL Unpublished
2 (bases 1 to 1768)
AUTHORS Wang,L., Willson,P., Chow,B., Gibbons,E. and Babiuk,L.
TITLE Direct Submission
JOURNAL Submitted (26-AUG-1998) VIDO, University of Saskatchewan, 120
Veterinary Road, Saskatoon, Saskatchewan S7N 5E3, Canada
FEATURES
source
1..1768
/organism="porcine circovirus"
/strain="B9"
/db_xref="taxon:46221"
51..995
/codon_start=1
/product="P35.8"
/protein_id="AAC61737.1"
/db_xref="PI:3661516"
/db_xref="GI:3661516"
/translation="MPSKNGRSGPQPHKRWVFTLNPNSEDERKKIRELPISLFDYFI
VGFEENEGRFTPLQGFANFVKOTFNKVLGARCHIEKAKGTDOONKEYCSKEGN
LLTECAPRSQGRSDLSAVSILLESGLVTVAKQHPVTFVKFRGLAELLKVSQGM

```

```

QKRDWKNVHFIVGPPGCKSKWAANFANPETTYWPKPKKMWDCYHGEKVVVDDFY
GMLPDDLLRLCDRYPLTVKTKGTVPFLARSILITSNOTPLEWSSSTAVPAVEALYR
RITSIVFWKNVTEQSTEGGQFVLSPPCFEPEINY"
complement(1034..1735)
/codon_start=1
/product="P27.9"
/protein_id="AAC61738.1"
/db_xref="PI:3661517"
/db_xref="GI:3661517"
/translation="MTYPRRRYRRRRHPRSHLGOILRRRRLVHPRHRYRWRKNGI
FNRLSRITFGYTVKATVTTPSWADMMRFNIDFVPPGGTNKISIFEYIRKRV
VEFWPCSPITQCDRGVSTAVILDNFVKATALTATYDYNYSRSHHTTQPFPFSHSRY
FTPKPVLDSITDYFPNNKRNQLWLRLQTSRNVDBHVLGTAFENSKYDODYNIRVTM
VQFRENLKDPPLKP"
BASE COUNT 463 a 357 c 482 g 466 t
ORIGIN
alignment_scores:
Quality: 1257.00 Length: 233
Ratio: 5.442 Gaps: 0
Percent Similarity: 99.142 Percent Identity: 98.283
alignment_block:
US-09-209-961-5 x AF086834/rev ..
Align seg 1/1 to reverse of: AF086834 from: 1 to: 1768
1 MetThrTyrProArgArgTyrArgArgArgHisArgProArgse 17
1735 ATGACGTATCCAGGAGGGGTTACCGCAGAGAAGACACCGCCCGCGAG 1686
17 rHisLeuGlyGlnIleLeuArgArgProTrpLeuValHisProArgH 34
1685 CCATCTTGCCAGATCTCGCGCGCCCTGGCTCGTCACCCCGCGCC 1636
34 isArgTyrArgTyrArgArgLysAsnGlyIlePheAsnThrArgLeuSer 50
1635 ACCGCTACCGTTGGAGAAGAAAATGGCATCTTCAACACCCCGCTCTCC 1586
51 ArgThrPheGlyTyrThrValLysArgTyrThrValThrThrProSerTr 67
1585 CGCACCTTCGGATATACTGTCAAGGCTACACAGTCACACAGCCCTCTG 1536
67 palValAspMetMetArgPheLysIleAspPheValProGlyG 84
1535 GCGGTGGACATGATGAGATTAAATATGACGACTTTGTTCCCGCGGAG 1486
84 1yGlyThrAsnLysIleSerIleProPheGluTyrTyrArgIleArgLys 100
1485 GGGGACCAACAATAATCTATACCTTTGAATACACAGATAAGAAAG 1436
101 ValLysValGluPheTyrProCysSerProIleThrGlnGlyAspArgG1 117
1435 GTTAAGTTGAATCTGCGCTGCTCCCATCACCCAGGGGTAGGGG 1386
117 yValGlySerThrAlaValIleLeuAspAspAsnPheValThrLysAlat 134
1385 AGTGGCTCCACTGCTGTTATTTAGATGATAACTTTGTAACAAAGGCCA 1336
134 hrAlaLeuThrTyrAspProTyrValAsnTyrSerSerArgHisThrIle 150
1335 CAGCCCTAACCTATGACCCATATGTAACACTACTCTCCCGCCATACAATC 1286
151 ProGlnProPheSerTyrHisSerArgTyrPheThrProLysProValle 167
1285 CCCCACCTTCTCTACCACTCCCGTTACTTACACCCAAACCTGTCT 1236
167 uAspSerThrIleAspTyrPheGlnProAsnAsnLysArgAsnGlnLeu 184
1235 TGACTCCACTATTGATTACTTCCAAACCAATAAACAAGGAATCAGCTTT 1186
184 rPLeuArgLeuGlnThrSerGlyAsnValAspHisValGlyLeuGlyThr 200

```

```
1185 GCGTGGGCTACAAACCTTAGAAATGTGACACACGTAGGCTCGGCAC 1136
201 AlapheGluAsnSerLysTyrAspGlnAspTyrAsnIleArgValThrMe 217
1135 GCGTTGCGAAACAGTAATACGACACGAGGACTACAATATCGGTAAACCAT 1086
217 tTyrValGlnPheArgGluPheAsnLeuLysAspProProLeuGluPro 233
1085 GATATGACAAATTCAGAGAAATTAATCTTAAGACCCCCCACTTAACCC 1037
```

seq\_name: gb\_vi:PCAJ3185

```
seq_documentation_block:
LOCUS PCAJ3185 1768 bp DNA circular VRL 03-JUL-1998
DEFINITION Porcine circovirus DNA, complete genome, isolate ISU-31.
ACCESSION AJ223185
NID 93293040
VERSION AJ223185.1 GI:3293040
KEYWORDS complete genome.
SOURCE porcine circovirus.
ORGANISM porcine circovirus.
REFERENCE 1 (bases 1 to 1768)
AUTHORS Morozov, I.
TITLE Direct Submission
JOURNAL Submitted (02-JAN-1998) Morozov I., Veterinary Medical Research
Institute, Iowa State University, 1802 Elwood Dr. VMRI Ames, Iowa
50011, U.S.A
REFERENCE 2 (bases 1 to 1768)
AUTHORS Morozov, I., Sirinarumit, T., Sorden, S.D., Halbur, P.G., Morgan, M.K.,
Yoon, K.J. and Paul, P.S.
TITLE Detection of a novel strain of Porcine circovirus in pigs with
Postweaning Multisystemic Wasting Syndrome (PMWS)
JOURNAL Unpublished
FEATURES
source 1. 1768
/organism="porcine circovirus"
/virion
/isolate="ISU-31"
/specific_host="Sus scrofa"
/db_xref="taxon:46221"
51..995
/note="ORF1"
/codon_start=1
/protein_id="CAA11157.1"
/db_xref="PID:e1310034"
/db_xref="PID:g3293041"
/db_xref="GI:3293041"
/translation="MPKKNKRSQPQPKRWVFLNPNSEDERKKIRELPISLEDFYI
VGEENEGRTPHLGQFANFKVKTFFNKKVYLGARHIEKAKGDDQNKKEYCSKEGN
LLIECGAPRSQGSRLSTAVSLTLESGSLVTVAEQHPVTFVRNFRGLAEILKVSQKM
QKRWKTNVHVIVGPGCGSKWAANFADPETTYKPPNKNWDGHHGEVYVDDFY
GWLPPDLLRLCDRYPLTVETKGTVPFLARSILITSNOTPLEWYSSSTAVPAVEALYR
RISLVFKNKAQTEGSEGGQFVLSPPCPPEFYNI"
```

CDS

```
BASE COUNT 452 a 361 c 492 g 463 t
ORIGIN
alignment_scores:
Quality: 1251.00 Length: 233
Ratio: 5.439 Gaps: 0
Percent Similarity: 98.712 Percent Identity: 97.854
```

alignment\_block:

US-09-209-961-5 x PCAJ3185/rev ..

Align seg 1/1 to reverse of: PCAJ3185 from: 1 to: 1768

```
1 MetThrTyrProArgArgTyrArgArgArgHisArgProArgSe 17
1735 ATGACGATATCCAGGAGGCGTTACCCGAGNAGACACCCGCCCGCAG 1686
```

```
17 rHisLeuGlyGlnIleLeuArgArgArgProTrpLeuValHisProArgH 34
1685 CCATCTTTGGCCAGATCTCCGCGCGCCCTGGCTGCTCCACCCCGCC 1636
34 isArgTyrArgTyrArgArgLysAsnGlyIlePheAsnThrArgLeuSer 50
1635 ACCCTACCGTTGGAGAGGAAAAATGCGATCTTCAACACCCGCTCTCC 1586
51 ArgThrPheGlyTyrThrValLysArgThrThrValThrThrProSerTr 67
1585 CGCACCTTCGGATATCTGTCAAGGCTACCACAGTCAGACGCCGCTCTG 1536
67 palValAspMetMetArgPheLysIleAspAspPheValProProGlyG 84
1535 GCGCGTGGACATGATGAGATTATATATGAGACTTTGTTCCCGCGGAG 1486
84 lyGlyThrAsnLysIleSerIleProPheGluTyrTyrArgIleArgLys 100
1485 GGGGACCAACAAATCTCTATACCTTTGAATACTACAGAAATAAGAAAG 1436
101 ValLysValGluPheThrProCysSerProIleThrGlnGlyAspArgGl 117
1435 GTTAAGGTTGAATTTCTGGCCCTGCTCCCCCATCACCCAGGCTGATAGGG 1386
117 yValGlySerThrAlaValIleLeuAspAspAsnPheValThrLysAlat 134
1385 AGTGGGCTCCACTGCTCTTATCTAGATGATGATGATGATGATGATGAT 1336
134 hrAlaLeuThrTyrAspProTyrValAsnTyrSerSerArgHisThrIle 150
1335 CAGCCCTAACCTATGACCATATGTAATACTACTCTCCCGGCATACAATC 1286
151 ProGlnProPheSerTyrHisSerArgTyrPheThrProLysProValLe 167
1285 CCCCAACCTTCTCTCCACCTCCCGTTACTTCACACCCCAACCTGTCT 1236
167 uAspSerThrIleAspTyrPheGlnProAsnAsnLysArgAsnGlnLeut 184
1235 TGACTCCACTATTGATGATGATGATGATGATGATGATGATGATGATGAT 1186
184 rpLeuArgLeuGlnThrSerGlyAsnValAspHisValGlyLeuGlyThr 200
1185 GGCTAAGGCTACAAACCTTAGAAATGTGGACACGCTAGGCTCGGCAC 1136
201 AlapheGluAsnSerLysTyrAspGlnAspTyrAsnIleArgValThrMe 217
1135 GCGTTGCGAAACAGTAATACGACACGAGGACTACAATATCGGTAAACCAT 1086
217 tTyrValGlnPheArgGluPheAsnLeuLysAspProProLeuGluPro 233
1085 GATATGACAAATTCAGAGAAATTAATCTTAAGACCCCCCACTTAACCC 1037
```

seq\_name: gb\_vi:AF166528

```
seq_documentation_block:
LOCUS AF166528 1768 bp mRNA VRL 19-AUG-1999
DEFINITION Porcine circovirus complete genome.
ACCESSION AF166528
NID 95739338
VERSION AF166528.1 GI:57393338
KEYWORDS porcine circovirus.
SOURCE porcine circovirus
ORGANISM porcine circovirus
REFERENCE 1 (bases 1 to 1768)
AUTHORS Yang, K.H., Lee, Y.F., Chao, D.S., Shieh, Y.C. and Lai, S.S.
TITLE Complete nucleotide sequences of porcine circovirus Tainan strand
outbreak in Taiwan
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1768)
AUTHORS Yang, K.H., Lee, Y.F., Chao, D.S., Shieh, Y.C. and Lai, S.S.
TITLE Direct Submission
JOURNAL Submitted (08-JUL-1999) Veterinary Medicine, National Chia-Yi
```

Institute of Technology, 300 Shei Fu Road, Lu Liao Li, Chia-yi City  
600, Taiwan

## FEATURES

Location/Qualifiers

1..1768  
/organism="porcine circovirus"  
/virion  
/strain="Tainan"  
/db\_xref="taxon:46221"  
/country="Taiwan"  
51..995

## CDS

/codon\_start=1  
/product="putative Rep protein"  
/protein\_id="AA050432.1"  
/db\_xref="PID:g5739339"  
/db\_xref="GI:5739339"  
/translation="MPSKNGRSQPQPKRWVFLNPNSEDERKKIRELPISLFDYFI  
VGEENEGRTPLHGFANFKVKKQTFNFKVGFARGHIEKAKGTDOQNKKEYCSKEGN  
LLIECAPRSQGSQSDLSLSTAVSLLESGSLVTVAEQHPVTFVRNFRGLAELLKVSQGM  
OKRDKNTNHHVIGPPCGSKWAANFADPETTYKPKRNKWDGYHGEVVDYF  
GWLPPDLLRLCDRYPLTVETKGTVPFLARSILITSNOTPLEWYSSTAVPAVEALYR  
RITSLVFNKNATEQSTEGSGQFVTLSPCPPEPEYINI"

BASE COUNT 453 a 367 c 492 g 456 t  
ORIGIN

## alignment\_scores:

Quality: 1244.00 Length: 233  
Ratio: 5.385 Gaps: 0  
Percent Similarity: 99.142 Percent Identity: 96.996

## alignment\_block:

US-09-209-961-5 x AF166528/rev ..

Align seg 1/1 to reverse of: AF166528 from: 1 to: 1768

```
1 MetThrTyrProArgArgTyrArgArgArgGHisArgProArgSe 17
|||||
1735 ATGCGTATCCAGGAGGGCTTCCGACAGCAAGACCGCCCGCGAG 1686
|||||
17 rHisLeuGlyGlnIleLeuArgArgProTrpLeuValHisProArgH 34
|||||
1685 CCATCTTGCCAGATCCTCCGCGCGCCCTGGCTCGCTCCACCCCGCC 1636
|||||
34 isArgTyrArgTrpArgArgLysAsnGlyIlePheAsnThrArgLeuSer 50
|||||
1635 ACCGTTACCGTGGAGAGAAATGGCATCTTCAACACCGCGCTCTCC 1586
|||||
51 ArgThrPheGlyTyrThrValLysArgThrThrValThrProSerTr 67
|||||
1585 CGACCTTCGGATATAGTCAAGGCTACACAGTCAGAACCGCCCTCTG 1536
|||||
67 pAlaValAspMetMetArgPheLysIleAspPheValProProGlyG 84
|||||
1535 GCGGTGGCATGATGAGATTATTAATTAACGACITTTGTTCCCGGGAG 1486
|||||
84 lyGlyThrAsnLysIleSerIleProPheGluTyrTyrArgIleArgLys 100
|||||
1485 GGGGACCAACAAAATCTATACCTTTGAATCTACAGATAAGAAAG 1436
|||||
101 ValLysValGluPheTrpProCysSerProIleThrGlnGlyAspArgG 117
|||||
1435 GTTAAGTTGAATTCGGCCCTGTTCCTCCCAATCACCGGGTGACAGGG 1386
|||||
117 yValGlySerThrAlaValIleLeuAspAspAsnPheValThrLysAla 134
|||||
1385 AGTGGCTCCACTGCTGTTATCTAGATGATAACITTTGTAAGTAAAGCCA 1336
|||||
134 hrAlaLeuThrTyrAspProTyrValAsnTyrSerSerArgHisThrIle 150
|||||
1335 CAGCCCTAACCTATGACCGGTATGAACACTACTCTCCCGCCATAACAATC 1286
|||||
151 ProGlnProPheSerTyrHisSerArgTyrPheThrProLysProValle 167
|||||
```

```
1285 CCCCAACCCCTTCTCTACCACTCCCGGTACTTTACCCCAACCTGTCT 1236
167 uAspSerThrIleAspTyrPheGlnProAsnAsnLysArgAsnGlnLeuT 184
|||||
1235 TGATCCACTATTGATTACTTCCACCAACAACAAGGAATCAGCTTT 1186
|||||
184 rPLeuArgLeuGlnThrSerGlyAsnValAspHisValGlyLeuGlyThr 200
|||||
1185 GCGTGAGGCTACAACTCGGCAATGTGGACCACCGTAGCGCTGGGCAC 1136
|||||
201 AlaPheGluAsnSerLysTyrAspGlnAspTyrAsnIleArgValThrMe 217
|||||
1135 GCGTTCCGAACAGTAATACGACGAGGACTACATATCGTGTAACTAT 1086
|||||
217 tTyrValGlnPheArgGluPheAsnLeuLysAspProProLeuGluPro 233
|||||
1085 GTATGTACAATTCAGAGAATTTAATCTTAAAGACCCCACTTAAACCC 1037
```

seq\_name: gb\_v1:AF027217

seq\_documentation\_block:

LOCUS AF027217 1768 bp DNA circular VRL 14-MAY-1998  
DEFINITION Porcine circovirus strain pmws PCV, complete genome.  
ACCESSION AF027217  
NID 92689645  
VERSION AF027217.1 GI:2689645

## KEYWORDS

SOURCE

porcine circovirus.

## ORGANISM

Viruses: ssDNA viruses; Circoviridae; Circovirus.

## REFERENCE

1 (bases 1 to 1768)

AUTHORS Hamel,A.L., Lin,L.L. and Nayar,G.P.

TITLE Nucleotide sequence of porcine circovirus associated with

postweaning multisystemic wasting syndrome in pigs

J. Virol. 72 (6), 5262-5267 (1998)

## MEDLINE

98241772

## REFERENCE

2 (bases 1 to 1768)

AUTHORS Hamel,A.L., Lin,L.L. and Nayar,G.P.S.

TITLE Direct Submission

JOURNAL Submitted (26-SEP-1997) Virology Laboratory, Veterinary Services

Branch, Manitoba Agriculture, 545 University Crescent, Winnipeg,

Manitoba R3T 5S6, Canada

## FEATURES

source

1..1768

/organism="porcine circovirus"

/strain="pmws PCV"

/db\_xref="taxon:46221"

/note="both strands of seven overlapping PCR fragments

were sequenced; virus isolated from lung, lymph node,

spleen and tonsil tissue from pigs affected by post

weaning multisystemic wasting syndrome"

51..995

/note="ORF1; similar to Rep protein encoded by

non-pathogenic PCV, Genbank Accession Number U49186;

predicted 35.8 kDa protein"

/codon\_start=1

/product="putative Rep protein"

/protein\_id="AAC59462.1"

/db\_xref="PID:g2689646"

/db\_xref="GI:2689646"

/translation="MPSKNGRSQPQPKRWVFLNPNSEDERKKIRELPISLFDYFI

VGEENEGRTPLHGFANFKVKKQTFNFKVGFARGHIEKAKGTDOQNKKEYCSKEGN

LLIECAPRSQGSQSDLSLSTAVSLLESGSLVTVAEQHPVTFVRNFRGLAELLKVSQGM

OKRDKNTNHHVIGPPCGSKWAANFADPETTYKPKRNKWDGYHGEVVDYF

GWLPPDLLRLCDRYPLTVETKGTVPFLARSILITSNOTPLEWYSSTAVPAVEALYR

RITSLVFNKNATEQSTEGSGQFVTLSPCPPEPEYINI"

117..125

/note="glycosylation site"

327..332

complement(357..671)

/note="ORF3; predicted 11.9 kDa protein"

/codon\_start=1

/protein\_id="AAC59464.1"

```

/db_xref="PID:g2689648"
/db_xref="GI:2689648"
/translation="MTVPLVSWFPCVCKISSPFAFTTPRWPHNDVYIGLPI
TLHFAHFKFQSPALSKRYKRVLLCNGHQTALQOQTHSSRQVPLSLRSSTF
NK"
CDS
complement(386..565)
/notes="ORF4; predicted 6.5 kDa protein"
/codon_start=1
/protein_id="AAC59465.1"
/db_xref="PID:g2689649"
/db_xref="GI:2689649"
/translation="MTCVLVQSRFCIFPLTFKKSAPRKFLTNVTCGSATVTRLP
SNKVTADVDSLRCP"
misc_feature
complement(470..478)
/notes="glycosylation site"
complement(688..753)
/notes="ORF8; predicted 2.3 kDa protein"
/codon_start=1
/protein_id="AAC59469.1"
/db_xref="PID:g2689653"
/db_xref="GI:2689653"
/translation="MDIDHIVSDVDPHTAASHKSHQ"
misc_feature
816..824
/notes="glycosylation site"
906..914
/notes="glycosylation site"
983..988
complement(989..1033)
/notes="ORF11; predicted 1.8 kDa protein"
/codon_start=1
/protein_id="AAC59472.1"
/db_xref="PID:g2689656"
/db_xref="GI:2689656"
/translation="MKNNHYEVIKTKQ"
polyA_signal
1016..1177
CDS
/notes="ORF5; predicted 6.2 kDa protein"
/codon_start=1
/protein_id="AAC59466.1"
/db_xref="PID:g2689650"
/db_xref="GI:2689650"
/translation="MVFIFHLGKVGKIKFSELYIHGYTDIVLVVYVTFERSAE
YVHISRGL"
polyA_site
complement(1022..1027)
complement(1034..1735)
/notes="ORF2; predicted 27.8 kDa protein"
/codon_start=1
/protein_id="AAC59463.1"
/db_xref="PID:g2689647"
/db_xref="GI:2689647"
/translation="MTYPRRRRRRRHRRSHLQGLRRRPLVHPHRYWRKNGI
FNRLSRFTGYKATVTPSPWADMMRFNIDFVPPGGTNKISIPFEYIRKRV
VEWPCSPITQGGRGVSTAVILDNDFVTATALTDPYVYNSRHTIPQPFYSHSR
FTKPVLDSTIDYFQPNKRTQLWLRLQTSRNVDRVGLGTAFENSIYDQYNIRVTY
VQREFNLKDPPLKP"
misc_feature
complement(1301..1309)
/notes="glycosylation site"
complement(1522..1611)
/notes="ORF6; predicted 3.1 kDa protein"
/codon_start=1
/protein_id="AAC59467.1"
/db_xref="PID:g2689651"
/db_xref="GI:2689651"
/translation="MASSTPASPAPSDILSRLOSERPPGRWT"
CDS
1524..1631
/notes="ORF10; predicted 4.1 kDa protein"
/codon_start=1
/protein_id="AAC59471.1"
/db_xref="PID:g2689655"
/db_xref="GI:2689655"
/translation="MSTAQEGVLTVALTVYPKVRRVRVLRVLRVLRVLRVLRV
1682..1741
/notes="ORF7; predicted 1.9 kDa protein"
/codon_start=1

```

```

/protein_id="AAC59468.1"
/db_xref="PID:g2689652"
/db_xref="GI:2689652"
/translation="MAAGAVSSSAVTPPWIRHS"
complement(join(1732..1768,1..92))
/notes="ORF9; predicted 4.6 kDa protein"
/codon_start=1
/protein_id="AAC59470.1"
/db_xref="PID:g2689654"
/db_xref="GI:2689654"
/translation="MWLGSSAILLAGHVAAEVLPRCCRSALVILTAHFFRFOL"
join(1746..1768,1..113)
/function="putative replication site"
join(1762..1768,1..2)
/notes="AAGTATTAC" is similar to the nonanucleotide motif
in the non-pathogenic PCV, GenBank Accession Number
U49186"
BASE COUNT 452 a 360 c 495 g 461 t
ORIGIN

alignment_scores:
  Quality: 1237.00      Length: 233
  Ratio: 5.402          Gaps: 0
  Percent Similarity: 98.283  Percent Identity: 96.996

alignment_block:
US-09-209-961-5 x AF027217/rev ..
Align seg 1/1 to reverse of: AF027217 from: 1 to: 1768

1 MetThrTyrProArgArgArgTyrArgArgArgArgHisArgProArgSe 17
1735 ATGACGCTATCCAAAGGAGCGGTACCGCAGAAGAACACCGCCCGCCGAG 1686

17 rHisLeuGlyGlnIleLeuArgArgArgProTyrLeuValHisProArgH 34
1685 CCACTCTGGCCAGATCTCCGCCCGCCCTGGCTGCTGCACCCCGCCGCC 1636

34 isArgTyrArgTyrArgArgLysAsnGlyIlePheAsnThrArgLeuSer 50
1635 ACCGCTACCGCTGGAGAAGAAATGGCATCTTCAACACCGCCCTCTCC 1586

51 ArgThrPheGlyTyrThrValLysArgThrThrValThrThrProSerTr 67
1585 CGCACCTTCGGATATATCTGTCAAGGCTACACAGTCAGAACGCCCTCTG 1536

67 pAlaValAspMetMetArgPheLysIleAspAspPheValProGlyG 84
1535 GCGCGTGGACATGATGAGATTATATATTGACGACTTTGTTCCCGCGGAG 1486

84 lyGlyThrAsnLysIleSerIleProPheGluTyrTyrArgIleArgLys 100
1485 GGGGACCACAAATCTCTATACCTTTGAATACTACAGATAAAGAAG 1436

101 ValLysValGluPheThrProCysSerProIleThrGlnGlyAspArgL 117
1435 GTTAAGGTTGAATCTGGCCCTGCTCCCATCATCACCCAGGCTGATAGGG 1386

117 yValGlySerThrAlaValIleLeuAspAspAsnPheValThrLysAla 134
1385 AGTGGGCTCCACTGCTCTTATTCTAGATGATACTTTGTAAACAAGCCA 1336

134 hrAlaLeuThrTyrAspProTyrValAsnTyrSerSerArgHisThrIle 150
1335 CAGCCCTAACCTATGACCCATATGTAACACTCTCTCCGCCATACATC 1286

151 ProGlnProPheSerTyrHisSerArgTyrPheThrProLysProValLe 167
1285 CCCCAACCTTCTCTACCACTCCCGTTACTTTCACACCCAAACCTGTCT 1236

167 uAspSerThrIleAspTyrPheGlnProAsnAsnLysArgAsnGlnLeuT 184
|||||

```

```

1235 TGACTCCACTATTGATTACTTCCCAACCAATAACAAAGGACTCAGCTTT 1186
184 rPLeuArgLeuGlnThrSerGlyAsnValaspHisValGlyLeuGlyThr 200
|||||
1185 GGCTGAGGCTACAACCTCTAGAATCTGGACACGAGTAGGCTCGGCACT 1136
|||||
201 AlaPheGluAsnSerLysTyrAspGlnAspTyrAsnIleArgValThrMe 217
|||||
1135 CGGTTCGAACACATATATACACGAGCTACAAATATCCGTGTAAACCAT 1086
|||||
217 tTyTValGlnPheArgGluPheAsnLeuLysAspProProLeuGluPro 233
|||||
1085 GTATGTACAATTCAGAGAATTATCTTAAAGACCCCTTAACCC 1037
|||||

seq_name: gb_vi:AF112862

seq_documentation_block:
LOCUS AF112862 1768 bp DNA circular VRL 06-JAN-1999
DEFINITION Porcine circovirus type 2-B, complete genome.
ACCESSION AF112862
NID 94106915
VERSION AF112862.1 GI:4106915
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 1768)
AUTHORS Hamel,A.L. and Nayar,G.P.S.
TITLE Nucleotide sequence of four different isolates of porcine
circuitvirus detected in pigs with various clinical syndromes
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 1768)
AUTHORS Hamel,A.L. and Nayar,G.P.S.
TITLE Direct Submission
JOURNAL Submitted (10-DEC-1998) Virology Laboratory, Manitoba Agriculture,
545 University Crescent, Winnipeg, Manitoba R3T 5S6, Canada
FEATURES
Location/Qualifiers
1..1768
/organism="porcine circovirus type 2-B"
/db_xref="taxon:85709"
/notes="sequence obtained from several overlapping PCRs
using DNA extracted from lung, mesenteric lymph node and
tonsil of pig; similar to Porcine circovirus sequence
presented in Genbank Accession Number AF027217; type-B
designation is based upon restriction endonuclease
digestion pattern"
13..35
repeat_region /rpt_type=tandem
/rpt_unit=13..18
CDS
51..995
/notes="ORF-1"
/codon_start=1
/product="putative Rep and coat protein"
/protein_id="AAD03086.1"
/db_xref="PID:g4106918"
/db_xref="GI:4106918"
/translation="MPSRKNGSGPQPHKRWFTLNNSSEDEKREKIRELPISLFDYFI
VGEENEGRTPLHGFANFKQTFNKVYLGARCHEKAKGTQONKEYGSKGN
LLIECAPRSQSGORSLSTAVSTLLESGLVTVAEQHPVTFVNRFGLAELLKVSQGM
QKRWKTNVHVLVGGPCGCKSKWAANFADPTTYKPPRNKWDGYHGEFVVVDIFY
GWLPHDILLRLCDRYPILTVEYTKGTVPFLLARSILLISNQPLEWYSTAVPAVEALYR
RTLSLFWRNATEQSTEEGGQFVTLSPPCPEFFPEYNI"
327..332
polyA_signal
CDS
complement(357..671)
/codon_start=1
/product="ORF-3"
/protein_id="AAD03087.1"
/db_xref="PID:g4106919"
/db_xref="GI:4106919"
/translation="MVTIPLVSRWFPVCGFRVCKYKISSPFAFTTPRPHNDVIGLPI
TLLHFFAHQKFSQPAEISDKRYRVLNCHQTPALQQGTHSSRQVTPLSLRSSTF
NK"
complement(386..565)

```

```
rep_origin      join(1762..1768,1..2)
/Note="similar to the nonanucleotide motif of Porcine
circovirus presented in GenBank Accession Numbers AF027217
and U49186"
BASE COUNT      448 a 360 c 493 g 467 t
ORIGIN

alignment_scores:
  Quality: 1237.00      Length: 233
  Ratio: 5.402          Gaps: 0
  Percent Similarity: 98.283 Percent Identity: 96.567

alignment_block:
US-09-209-961-5 x AF112862/rev ..
Align seg 1/1 to reverse of: AF112862 from: 1 to: 1768
1 MetThrTyrProArgArgArgTyrArgArgArgHisArgProArgSe 17
1735 ATGACGTATCCAGGAGCGGTACCGAGAAGACACACCGCCCGCAG 1686
17 rHisLeuGlyGlnIleLeuArgArgProTrpLeuValHisProArgH 34
1685 CCATCTTGGCAGATCTCCGCCCGCCCTGGCTCGCTCCACCCGCC 1636
34 isArgTyrArgTTPArgArgLysAsnGlyIlePheAsnThrArgLeuSer 50
1635 ACCGTACCGTGGAGGAAGAAATGCACTTCAACACCGCCCTCC 1586
51 ArgThrPheGlyTyrThrValLysArgThrThrValThrProSerTr 67
1585 CGCACCTTCGATATACGTCAAGGCTACACACAGTCACACGCCCTCTG 1536
67 pAlaValAspMetMetArgPheLysIleAspPheValProProGlyG 84
1535 GCGCGGTGCAGATGATGATTAATATGACGACTTGTTCGCCCGGAG 1486
84 lYgLYThrAsnLysIleSerIleProPheGluTyrTyrArgIleArgLys 100
1485 GGGGGACCAAAATCTGTGATCCCTTGAATCTACTACAGAATAAGAAG 1436
101 ValLysValGluPheTTPProCysSerProIleThrGlnGlyAspArgG 117
1435 GTTAAGTTGAATCTCGCCCTGCTCCCATACCCAGGTGATAGGG 1386
117 yValGlySerThrAlaValIleLeuAspAspAsnPheValThrLysAla 134
1385 AGTGGGCTCCACTGCTTATTCTAGATGATACTTTGTAACAAAGGCCA 1336
134 hrAlaLeuThrTyrAspProTyrValAsnTyrSerSerArgHisThrIle 150
1335 CAGCCCTAACCTATGACCCATATGAACTACTCTCCCGGCATACAATC 1286
151 ProGlnProPheSerTyrHisSerArgTyrPheThrProLysProValLe 167
1285 CCCCAACCCCTCTCTACCACTCCCGTTAATACACCAACCACTGTACT 1236
167 uAspSerThrIleAspTyrPheGlnProAsnAsnLysArgAsnGlnLeu 184
1235 TGACAGAACTATTGATTACTTCCAACTAATAACAAAAAATCAGCTTT 1186
184 rPLeuArgLeuGlnThrSerGlyAsnValAspHisValGlyLeuGlyThr 200
1185 GCGTGGGCTACAAACCTCTAGAAATGTGGACACGCTAGGCTCGGCAC 1136
201 AlaPheGluAsnSerLysTyrAspGlnAspTyrAsnIleArgValThrMe 217
1135 GCGTTCGAAACAGTAATAACGACGAGCTACATAATATCCGTGTGAAC 1086
217 tTyrValGlnPheArgGluPheAsnLeuLysAspProProLeuGluPro 233
1085 GTATGTACAATTCAGAAATTAATCTTAAGACCCCACTTAACCC 1037
```

seq\_name: gb\_vi:AF147751

seq\_documentation\_block: 1768 bp DNA VRL 08-JUN-1999  
LOCUS AF147751 Porcine circovirus type 2 unknown genes.  
DEFINITION AF147751  
ACCESSION  
NID 95007010

VERSION AF147751.1 GI:5007010

KEYWORDS porcine circovirus type 2.

SOURCE porcine circovirus type 2.

ORGANISM Viruses; ssDNA viruses; Circoviridae; Circovirus.

REFERENCE 1 (bases 1 to 1768)

AUTHORS Pogranichnyy, R., Yoon, K.-J., Harms, P., Swenson, S., Zimmerman, J. and Sorden, S.

TITLE Characterization of clinical and immune responses in young swine to experimental porcine circovirus type II infection.

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1768)

AUTHORS Yoon, K.-J. and Pogranichnyy, R.M.

TITLE Direct Submission

JOURNAL Submitted (30-APR-1999) Veterinary Diagnostic Laboratory, Iowa State University, College of Veterinary Medicine, Ames, Iowa 50011, USA

FEATURES Location/Qualifiers

1..1768

/organism="porcine circovirus type 2"

/isolate="ISUVDL 98-15237"

/specific\_host="Sus scrofa"

/db\_xref="taxon:85708"

/note="isolated from pigs with postweaning multisystemic wasting syndrome"

51..995

/note="ORF1; 35-36kd"

/codon\_start=1

/product="unknown"

/protein\_id="AAD37776.1"

/db\_xref="PID:g5007011"

/db\_xref="GI:5007011"

/translation="MPSKNGRSGPQPKRWFTLNNPSEDERKKIRELPISLDYFI

VGEENEGRTPHLOGFANFVKQTFNKVWYLGARCHIEKAKGTDOONKEYCSKEG

LLIECGAPRSQGRSGLSTAVSTLLESGSLVTAEOHPVTFVNRFLAETLLKYSKM

OKRDWRTNVIYVGPCCGSKNAANFADPTFYWKPRKNWDCYGEVEVVDYDFY

GWLPPDLLRLCDRIPLTVETKGTVPFLARSLITSNQTPLEWYSTAVPAVEALYR

RITSLVFWKNATEQSTEEGGQFVTLSPPCPEFFEYIN"

complement(1034..1735)

/note="ORF2; 27-28kd"

/codon\_start=1

/product="unknown"

/protein\_id="AAD37777.1"

/db\_xref="PID:g5007012"

/db\_xref="GI:5007012"

/translation="MTYPRRRYRRRRPRSHLQILRRRPLVHPRRYVRRRNGI

FNRLSTFGYTVKATVTRTPSWAVMMNRNIDDFVPPGGTNKISIPFEYRIRKVK

VEFWPCSPITGGDRGVSTAVILDNFVTKATALTIDPTVYNSRRHTIPQFSCHSRY

FTPQVLDISTYGFQPNKRNQLWLRLQTSRNVHDVGLGTAFENSIDYQDYNIRVTY

VQFREFNLKDPPLKP"

BASE COUNT 450 a 362 c 494 g 462 t

ORIGIN

alignment\_scores:

Quality: 1234.00 Length: 233

Ratio: 5.412 Gaps: 0

Percent Similarity: 97.854 Percent Identity: 96.996

alignment\_block:

US-09-209-961-5 x AF147751/rev ..

Align seg 1/1 to reverse of: AF147751 from: 1 to: 1768

1 MetThrTyrProArgArgArgTyrArgArgArgHisArgProArgSe 17

```

|||||
1735 ATGACGTATCAAGAGCGGTACCGCAGAGAGACACACCCCGCAG 1686
17 rHisLeuGlyGlnIleLeuArgArgProTrpLeuValHisProArgH 34
|||||
1685 CCATCTCGGCAGATCTCCGCGCGCCCTGGCTGCTCCACCCCGCC 1636
34 isArgTyArgTrpArgArgLysAsnGlyIlePheAsnThrArgLeuSer 50
|||||
1635 ACCGCTACCGCTGTGAGAAGGAATGATCTTCAACACCGCGCTGCC 1586
51 ArgThrPheGlyTyThrValLysArgThrThrValThrTrpSerTr 67
|||||
1585 CGACCTTCGGATATCTGTCACGGCTACACAGTCAGACGCCCTCATG 1536
67 pAlaValAspMetMetArgPheLysIleAspPheValProProGlyG 84
|||||
1535 GCGCGTGCATGATGAGATTATATGACGACTTCTTCCCGCGGAG 1486
84 lyGlyThrAsnLysIleSerIleProPheGlyTyThrArgIleArgLys 100
|||||
1485 GGGGGACCAACAAATATCCATACCTTTGAATATACAGAAATAGAAG 1436
101 ValLysValGluPheTTPProCysSerProIleThrGlnGlyAspArgG 117
|||||
1435 GTTAAGTTGAATCTGGCCCTGTCTCCCATCACCGAGGTATAGGGG 1386
117 yValGlySerThrAlaValIleLeuAspAspAsnPheValThrLysAla 134
|||||
1385 AGTGGCTCCACTGCTGTATTCTAGATGATAACTTTGTAACAAAGGCCA 1336
134 hrAlaLeuThrTyAspProTyValAsnTySerSerArgHisThrIle 150
|||||
1335 CAGCCCTAACTATGACCATATGATTAATCTCTCCCGCATACAATC 1286
151 ProGlnProPheSerTyHisSerArgTyThrPheProLysProValIle 167
|||||
1285 CCCCACCTTCTCTGCCACTCCGTTACTTCACACCCAAACCTGTCT 1236
167 uAspSerThrIleAspTyThrPheGlnProAsnAsnLysArgAsnGlnLeu 184
|||||
1235 TGACTCCACTATTGATTCTTCCACCAATAAACAAGGAATCAGCTTT 1186
184 rPLeuArgLeuGlnThrSerGlyAsnValAspHisValGlyLeuGlyThr 200
|||||
1185 GCCTGAGGCTACAAACCTCTAGAAATGTGGACCGTAGGCTCGGCAG 1136
201 AlaPheGluAsnSerLysTyAspGlnAspTyAsnIleArgValThrMe 217
|||||
1135 CGGTTCGAAACAGTATATACAGCAGACACTACAATATCCGTGTAAACAT 1086
217 tTyrValGlnPheArgGluPheAsnLeuLysAspProProLeuGluPro 233
|||||
1085 GTATGTACAATTACAGAGAATTAACTTTAAAGACCCCGCCCTTAAACC 1037
seq_name: gb_vi:AF109397
seq_documentation_block:
LOCUS AF109397 1768 bp DNA circular VRL 02-FEB-1999
DEFINITION Bovine circovirus, complete genome.
ACCESSION AF109397
NID 94210313
VERSION AF109397.1 GI:4210313
KEYWORDS
SOURCE
ORGANISM
bovine circovirus.
VIRUSES: ssDNA viruses; Circoviridae; Circovirus.
REFERENCE
1 (bases 1 to 1768)
AUTHORS
Hamel,A.L. and Nayar,G.P.S.
TITLE
Nucleotide sequence of a circovirus detected in cattle with various
clinical syndromes
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 1768)

```

```

AUTHORS
Hamel,A.L. and Nayar,G.P.S.
TITLE
Direct Submission
JOURNAL
Submitted (27-NOV-1998) Virology Laboratory, Veterinary Diagnostic
Laboratory, Manitoba Agriculture, 545 University Crescent,
Winnipeg, Manitoba R3T 5S6, Canada
REFERENCE
3 (bases 1 to 1768)
AUTHORS
Hamel,A.L. and Nayar,G.P.S.
TITLE
Direct Submission
JOURNAL
Submitted (02-FEB-1999) Virology Laboratory, Veterinary Diagnostic
Laboratory, Manitoba Agriculture, 545 University Crescent,
Winnipeg, Manitoba R3T 5S6, Canada
REMARK
Sequence update by submitter
COMMENT
On Feb 2, 1999 this sequence version replaced gi.4028609.
FEATURES
source
1..1768
/organism="bovine circovirus"
/virion
/db_xref="taxon:85542"
/notes="BCV
sequence was obtained from several overlapping PCRs using
DNA extracted from various tissues from cattle; similar to
Porcine circovirus sequence presented in GenBank Accession
Number AF027217"
13..35
repeat_region
/rpt_type=tandem
/rpt_unit=13..18
51..995
CDS
/notes="ORF-1; similar to Porcine circovirus ORF1 encoded
by the sequence presented in GenBank Accession Number
AF027217."
/codon_start=1
/product="putative Rep and coat protein"
/protein_id="AAD11928.1"
/db_xref="PID:g4210314"
/db_xref="GI:4210314"
/translation="MPSKNGRSGPOPHKRWFTLNPSDEKRRKIRELPSLFDYFI
VGEENEGRTPHLQGFANFKOTNKVKYLGARCHIIEKAGTDOQKEYCSKEGN
LLIECGAPRSGQSRDLSTAVSTLLSGSLTVAEQHPVTFVRNFRGLAELLKVSCKM
QKRDKNTNHHVIVGPPCGSKWAANFADPTTYKPRNKKWMDGYHGEVVVYIDDFY
GWLPPDLLRLCDRPLTVETGKGTVPFLARSILITSNQTPLEWYSSTAVPAVEALYR
RITSLVFNKNAEQSTEEGGQFVTLSPPCPEFFPEYNY"
327..332
polyA_signal
complement(357..671)
/notes="similar to Porcine circovirus ORF-3 encoded by the
sequence presented in GenBank Accession Number AF027217."
/codon_start=1
/product="ORF-3"
/protein_id="AAD11930.1"
/db_xref="PID:g4210315"
/db_xref="GI:4210315"
/translation="MVTIPPLVSRWPFVCGFRVCKISSPFAFTTPRPHNDVYIGLPI
TLHFFPAHFQKFSQPAEISDKYRVLLCNHQTPALQOQTHSSRQVTPLSRRSSTF
NK"
complement(386..565)
CDS
/notes="similar to Porcine circovirus ORF-4 encoded by the
sequence presented in GenBank Accession Number AF027217."
/codon_start=1
/product="ORF-4"
/protein_id="AAD11931.1"
/db_xref="PID:g4210316"
/db_xref="GI:4210316"
/translation="MCTLVFQSRFCIFPLTFKSSASPRKFLNVTGCCSATVTRLPL
SNKVLTAVDRSLRCP"
553..732
CDS
/codon_start=1
/product="ORF-12"
/protein_id="AAD11927.1"
/db_xref="PID:g4028610"
/db_xref="GI:4028610"
/translation="MYTSLMGHLGVKANGLLILQTRKPTGHNHLETSGGMYTMVKKW
LLMTFMAGCRGMIV"
complement(688..753)
CDS
/codon_start=1

```



Align seq 1/1 to reverse of: AFL09397 from: 1 to: 1768

```
REFERENCE 2 (bases 1 to 1768)
AUTHORS Kuo,T.Y., Chiou,Y.C. and Lai,S.S.
TITLE Direct Submission
JOURNAL Submitted (27-MAY-1999) Veterinary Medicine, National Taiwan
University, 142, Chousan Road, Taipei 106, Taiwan
FEATURES
source
1..1768
/organism="porcine circovirus"
/isolate="MLW98"
/db_xref="taxon:46221"
/country="Taiwan"
51..995
/notes="putative Rep protein"
/codon_start=1
/product="ORF1"
/protein_id="AAD38398.1"
/db_xref="PID:g5052005"
/db_xref="GI:5052005"
/translation="MPSKNGRSGPQPHKRWFTLNNPSEDERKKIRELPISLDFCFI
VGEENEGRTPLHQGFANFKVKTENKVKYFCARCHIEKAGTDOONKEYCSKGN
LLIECGAPRSQGSRLSTAVSLLESGLVTVAEQHPVTFVRNFRGLAELLKVSQGM
QKRDWKNVHVIVGPPCGSKWAANFADPETTYWKPRKNWMDGYHGEVWVDDFY
GWLPPDILLRLCDRYPLTVETKGTGTFVLARSILITSNQPLEWYSTAVPAVEALYR
RITSLVFNKNATEQSTEGGQFVTLSPPCPEFFPEINY"
BASE COUNT 451 a 367 c 495 g 455 t
ORIGIN

alignment_scores:
Quality: 1226.00 Length: 233
Ratio: 5.330 Gaps: 0
Percent Similarity: 98.712 Percent Identity: 95.708

alignment_block:
US-09-209-961-5 x AF154679/rev ..
Align seg 1/1 to reverse of: AF154679 from: 1 to: 1768
1 MetThrTyrProArgArgArgTyrArgArgArgArgHisArgProArgSe 17
|||||
1735 ATGAGTATCAAGGAGGCGTTTCGCGAGACGAAGACACCGCCCGCGAG 1686
|||||
17 rHisLeuGlyGlnIleLeuArgArgArgProTrpLeuValHisProArgH 34
|||||
1685 CCATCTTGCCAGATCCTCCGCGCGCGCCCTGCTGCTCCACCCCGCC 1636
|||||
34 isArgTyrArgTrpArgArgLysAsnGlyIlePheAsnThrArgLeuSer 50
|||||
1635 AGCGTTACCGCTGGAGAGGAAATGCGATCTTCAACACCGCGCTCTCC 1586
|||||
51 ArgThrPheGlyTyrThrValLysArgThrThrValThrProSerTr 67
|||||
1585 CGCACCTTCGGATATACGTCAAGGCTAGCACAGTACAGAACCGCTCTG 1536
|||||
67 pAlaValAspMetMetArgPheLysIleAspAspPheValProGlyG 84
|||||
1535 GCGGTGGACATGATGAGATTGTATTAAACGACTTGTTCCTCCCGGGAG 1486
|||||
84 lyGlyThrAsnLysIleSerIleProPheGluTyrTyrArgIleArgLys 100
|||||
1485 GGGGACCACAAATCTCTATACCTTTGATACCTACAGATAGAAAG 1436
|||||
101 ValLysValGluPheTrpProCysSerProIleThrGlnGlyAspArgG 117
|||||
1435 GTTAAGGTGAATTCGCGCTGCTCCCAATCACCCAGGGTGACAGGGG 1386
|||||
117 yValGlySerThrAlaValIleLeuAspAspAsnPheValThrLysAla 134
|||||
1385 AGTGGGCTCCACTGCTGTATTCTAGATGATAACTTGTGAATAGGCCA 1336
|||||
134 hrAlaLeuThrTyrAspProTyrValAsnTyrSerArgHisThrIle 150
|||||
1335 CAGCCCTAACCTATGACCCCTATGTAACCTACTCTCCCGCCATACATC 1286
```

```
151 ProGlnProPheSerTyrHisSerArgTyrPheThrProLysProValLe 167
|||||
1285 CCCCAACACCTTCTCTACCACTCCCGGTACTTTACCCCAACCTGTCTT 1236
|||||
167 uaspSerThrIleAspTyrPheGlnProAsnAsnLysArgAsnGlnLeu 184
|||||
1235 TGATTCACATATTGATTACTTCCACCAACAACAAGAAGATCAGCTTT 1186
|||||
184 rPLeuArgLeuGlnThrSerGlyAsnValAspHisValGlyLeuGlyThr 200
|||||
1185 GCGTGAAGCTACAAACCTCGCAATGTGGACACGCTAGGCTCGGCAC 1136
|||||
201 AlaPheGluAsnSerLysTyrAspGlnAspTyrAsnIleArgValThrMe 217
|||||
1135 GCGTTCGAACAGTAATAACACAGGACTACATATCCGTGTAACTAT 1086
|||||
217 tTyrValGlnPheArgGluPheAsnLeuLysAspProLeuGluPro 233
|||||
1085 GTATGTACATTCAGAGATTTAATCTTAAGACCCCACTTAAACCC 1037
|||||
seq_name: gb_vi:AF109399
seq_documentation_block:
LOCUS AF109399 1768 bp DNA circular VRL 06-JAN-1999
DEFINITION Porcine circovirus type 2-E, complete genome.
ACCESSION AF109399
NID 94106896
VERSION AF109399.1 GI:4106896
KEYWORDS
SOURCE
ORGANISM porcine circovirus type 2-E.
Viruses; ssDNA viruses; Circoviridae; Circovirus; #59#line
REFERENCE
1 (bases 1 to 1768)
AUTHORS Hamel,A.L. and Nayar,G.P.S.
TITLE Nucleotide sequence of four different isolates of circovirus
detected in pigs with various clinical syndromes
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1768)
AUTHORS Hamel,A.L. and Nayar,G.P.S.
TITLE Direct Submission
JOURNAL Submitted (27-NOV-1998) Virology Laboratory, Veterinary Diagnostic
Laboratory, Manitoba Agriculture, 545 University Crescent,
Winnipeg, Manitoba R3T 5S6, Canada
FEATURES
source
1..1768
/organism="porcine circovirus type 2-E"
/virus
/db_xref="taxon:85544"
/notes="similar to Porcine circovirus sequence presented in
GenBank Accession Number AF027217; type-E designation is
based upon restriction endonuclease digestion pattern;
sequence obtained from several overlapping PCRs using DNA
extracted from lung, mesenteric lymph node and tonsil of
pig"
repeat_region 13..35
/rpt_type=tandem
/rpt_unit=13..18
51..995
/notes="ORF-1"
/codon_start=1
/product="putative Rep and coat protein"
/protein_id="AAD03071.1"
/db_xref="PID:g4106897"
/db_xref="GI:4106897"
/translation="MPSKNGRSGPQPHKRWFTLNNPSEDERKKIRELPISLDFYI
VGEENEGRTPLHQGFANFKVKTENKVKYFCARCHIEKAGTDOONKEYCSKGN
LLIECGAPRSQGSRLSTAVSLLESGLVTVAEQHPVTFVRNFRGLAELLKVSQGM
QKRDWKNVHVIVGPPCGSKWAANFADPETTYWKPRKNWMDGYHGEVWVDDFY
GWLPPDILLRLCDRYPLTVETKGTGTFVLARSILITSNQPLEWYSTAVPAVEALYR
RITSLVFNKNATEQSTEGGQFVTLSPPCPEFFPEINY"
327..332
polyA_signal
```

```
CDS
complement(357..671)
/codon_start=1
/product="ORF-3"
/protein_id="AAD03073.1"
/db_xref="PID:g4106899"
/db_xref="GI:4106899"
/translation="MVTIPPLVSRWFPVCGFRVCKISSPFAFTTTPRPWPHNDVYIGLPI
TLHFPAHQFQSFQPAEISDKRYRLLONGHOTPALQOQTHSSRQVTPLSRSSTL
NK"
complement(386..565)
/codon_start=1
/product="ORF-4"
/protein_id="AAD03074.1"
/db_xref="PID:g4106900"
/db_xref="GI:4106900"
/translation="MTCTLVFQSRFCIFPLFKSASPRKFLNTVGCSTVTRPLP
SNKVLTAVDRLRCP"
553..732
/codon_start=1
/product="ORF-12"
/protein_id="AAD03081.1"
/db_xref="PID:g4106907"
/db_xref="GI:4106907"
/translation="MYSLWHLGVVVKANGLLILQTRKPHGTGNHLETSGGMVTVMKW
LLMTFMAGCRGMIV"
complement(688..753)
/codon_start=1
/product="ORF-8"
/protein_id="AAD03078.1"
/db_xref="PID:g4106904"
/db_xref="GI:4106904"
/translation="MDIDHTVSVDPHTAASHKSHQ"
983..988
complement(989..1033)
/codon_start=1
/product="ORF-11"
/protein_id="AAD03080.1"
/db_xref="PID:g4106906"
/db_xref="GI:4106906"
/translation="MNNKNHYEVKKTO"
1016..1363
/notes="similar to Porcine circovirus ORF-5 encoded by the
sequence presented in GenBank Accession Number AF027217."
/codon_start=1
/product="ORF-5"
/protein_id="AAD03075.1"
/db_xref="PID:g4106901"
/db_xref="GI:4106901"
/translation="MVFIIHLGFKWGVKIFSELYIHGYTDIVLVLVVTFVRSAAE
YVHIHSTGLSHPOLIPFVIVLEVINSGIKRFGCEVPVGVGEGIGDCMAGGVVYVG
IGLGCGLYYKKVI"
complement(1022..1027)
complement(1034..1735)
/codon_start=1
/product="ORF-2"
/protein_id="AAD03072.1"
/db_xref="PID:g4106898"
/db_xref="GI:4106898"
/translation="MTVPRRRRRRHRPRSHLQQLRRRPLWHPHRYRWKRNKI
FNARLSTFGTVATVSTFSAVMDLRFNLDVFPVCGGTGNKISIFEYRIKVK
VEFWPCSPITOGDGVSSAILLDNDFVKATQADPYVNYSSRHTIPQPFYSRY
FTRPVLDSITIDYFQPNKNRNLQMLQTSRNVHDVGLGTAFENSKYQDYNIRVTM
VQREFNLKDPPLKP"
1524..1631
/codon_start=1
/product="ORF-10"
/protein_id="AAD03079.1"
/db_xref="PID:g4106905"
/db_xref="GI:4106905"
/translation="MSTAQEGVLTVALTVTPKVRERALKMPFLFOR"
complement(1528..1611)
/codon_start=1
/product="ORF-6"
complement(357..671)
/codon_start=1
/product="ORF-3"
/protein_id="AAD03073.1"
/db_xref="PID:g4106899"
/db_xref="GI:4106899"
/translation="MVTIPPLVSRWFPVCGFRVCKISSPFAFTTTPRPWPHNDVYIGLPI
TLHFPAHQFQSFQPAEISDKRYRLLONGHOTPALQOQTHSSRQVTPLSRSSTL
NK"
complement(386..565)
/codon_start=1
/product="ORF-4"
/protein_id="AAD03074.1"
/db_xref="PID:g4106900"
/db_xref="GI:4106900"
/translation="MTCTLVFQSRFCIFPLFKSASPRKFLNTVGCSTVTRPLP
SNKVLTAVDRLRCP"
553..732
/codon_start=1
/product="ORF-12"
/protein_id="AAD03081.1"
/db_xref="PID:g4106907"
/db_xref="GI:4106907"
/translation="MYSLWHLGVVVKANGLLILQTRKPHGTGNHLETSGGMVTVMKW
LLMTFMAGCRGMIV"
complement(688..753)
/codon_start=1
/product="ORF-8"
/protein_id="AAD03078.1"
/db_xref="PID:g4106904"
/db_xref="GI:4106904"
/translation="MDIDHTVSVDPHTAASHKSHQ"
983..988
complement(989..1033)
/codon_start=1
/product="ORF-11"
/protein_id="AAD03080.1"
/db_xref="PID:g4106906"
/db_xref="GI:4106906"
/translation="MNNKNHYEVKKTO"
1016..1363
/notes="similar to Porcine circovirus ORF-5 encoded by the
sequence presented in GenBank Accession Number AF027217."
/codon_start=1
/product="ORF-5"
/protein_id="AAD03075.1"
/db_xref="PID:g4106901"
/db_xref="GI:4106901"
/translation="MVFIIHLGFKWGVKIFSELYIHGYTDIVLVLVVTFVRSAAE
YVHIHSTGLSHPOLIPFVIVLEVINSGIKRFGCEVPVGVGEGIGDCMAGGVVYVG
IGLGCGLYYKKVI"
complement(1022..1027)
complement(1034..1735)
/codon_start=1
/product="ORF-2"
/protein_id="AAD03072.1"
/db_xref="PID:g4106898"
/db_xref="GI:4106898"
/translation="MTVPRRRRRRHRPRSHLQQLRRRPLWHPHRYRWKRNKI
FNARLSTFGTVATVSTFSAVMDLRFNLDVFPVCGGTGNKISIFEYRIKVK
VEFWPCSPITOGDGVSSAILLDNDFVKATQADPYVNYSSRHTIPQPFYSRY
FTRPVLDSITIDYFQPNKNRNLQMLQTSRNVHDVGLGTAFENSKYQDYNIRVTM
VQREFNLKDPPLKP"
1524..1631
/codon_start=1
/product="ORF-10"
/protein_id="AAD03079.1"
/db_xref="PID:g4106905"
/db_xref="GI:4106905"
/translation="MSTAQEGVLTVALTVTPKVRERALKMPFLFOR"
complement(1528..1611)
/codon_start=1
/product="ORF-6"
CDS
protein_id="AAD03076.1"
/db_xref="PID:g4106902"
/db_xref="GI:4106902"
/translation="MASSTPASPAPSDILSRILPQSARPPGR"
1682..1741
/codon_start=1
/product="ORF-7"
/protein_id="AAD03077.1"
/db_xref="PID:g4106903"
/db_xref="GI:4106903"
/translation="MAAGAVSSAETPPWIRHS"
join(1750..1768,1..13)
join(1762..1768,1..2)
/notes="putative; similar to the nonnucleotide motif of
porcine circovirus presented in GenBank Accession Numbers
AF027217 and U49186"
BASE COUNT 448 a 362 c 495 g 463 t
ORIGIN
alignment_scores
Quality: 1217.00 Length: 233
Ratio: 5.314 Gaps: 0
Percent Similarity: 98.283 Percent Identity: 93.562
alignment_block
US-09-209-961-5 x AF109399/rev ..
Align seg 1/1 to reverse of: AF109399 from: 1 to: 1768
1 MetThrTyrProArgArgArgTyrArgArgArgArgHisArgProArgSe 17
|||||
1735 ATGACGTATCCAGGAGGCGCTTCGCCGCGCCGCTGCTGCTCCACCCCGCAG 1686
17 rHisLeuGlyGlnIleLeuArgArgArgProTyrLeuValHisProArgH 34
|||||
1685 CCATCTGGCCAGATCTCCGCCGCGCCGCTGCTGCTCCACCCCGC 1636
34 IsArgTyrArgTyrArgArgLysAsnGlyIlePheAsnThrArgLeuSer 50
|||||
1635 ACCGTTACCGCTGAAAAAGAAATGGCATCTTCAACGCCCGCTCTCC 1586
51 ArgThrPheGlyTyrThrValLysArgThrThrValThrThrProSerTr 67
|||||
1585 CGCACCTTCGGATATCTGTCAGAGCTACACAGCTCAGCAGCGCTCTCTG 1536
67 PAlaValAspMetMetArgPheLysIleAspPheValProProGlyG 84
|||||
1535 GCGGTAGACATGCTGAGATTAATCTTGACGACTTTGTTCCCGCAGGAG 1486
84 lYgLYThrAsnLysIleSerIleProPheGlyTyrTyrArgIleArgLys 100
|||||
1485 GGGGACCACAAATCTCTATACCTTTGAATACTACAGATAAGAAAG 1436
101 ValLysValGluPheThrProCysSerProIleThrGlnGlyAspArgGl 117
|||||
1435 GTTAAGGTTGAATCTGCGCTGCTCCCGATCACCAGGCGTGACAGGG 1386
117 yValGlySerThrAlaValIleLeuAspAspAsnPheValThrLysAla 134
|||||
1385 AGTTGGATCCAGTCTATTATCTTAGATGACAACTTTGTAATAAGGCCA 1336
134 hrAlaLeuThrTyrAspProTyrValAsnTyrSerSerArgHisThrIle 150
|||||
1335 CAGCCCAACCTATGACCCCTACGTAACACTCTCTCCCGCCATACATC 1286
151 ProGlnProPheSerTyrHisSerArgTyrPheThrProLysProValLe 167
|||||
1285 CCCCAACCTTCTCTACCATCTCCAGGCTACTTCACACCAACCTGTCT 1236
167 uAspSerThrIleAspTyrPheGlnProAsnAsnLysArgAsnGlnLeu 184
|||||
1235 TGATTCCACTATTGATTACTTCCAAACCAATAACAAAAGGAATCAGCTGT 1186
```

```
184 rpLeuArgLeuGlnThrSerGlyAsnValAspHisValGlyLeuGlyThr 200
|||||
1185 GGATGAGACTCCAAACCACTAGAAATGTGGACCACTAGGCTCGGCAC 1136
|||||
201 AlaPheGluAsnSerLysTyAspGlnAspTyAsnIleArgValThrMe 217
|||||
1135 CGGTGCGAAACAGTAATACGACGACTACATATCCGTGTAAACCAT 1086
|||||
217 tTyValGlnPheArgGluPheAsnLeuLysAspProLeuGluPro 233
|||||
1085 GTATGTACAATTACAGAGATTAATCTTAAGACCCCACTAAACCC 1037
|||||
seq_name: gb_vi:AF109398

seq_documentation_block:
LOCUS AF109398 1768 bp DNA circular VRL 06-JAN-1999
DEFINITION Porcine circovirus type 2-C, complete genome.
ACCESSION AF109398
NID 94106885
VERSION AF109398.1 GI:4106885
KEYWORDS
SOURCE
ORGANISM
VIRUSES: ssDNA viruses; Circoviridae; Circovirus; #2#ine
circular type 2
1 (bases 1 to 1768)
Hamel, A.L. and Nayar, G.P.S.
Nucleotide sequence of four different isolates of circovirus
detected in pigs with various clinical syndromes
Unpublished
JOURNAL
REFERENCE
2 (bases 1 to 1768)
Hamel, A.L. and Nayar, G.P.S.
Direct Submission
TITLE
Submitted (27-NOV-1998) Virology Laboratory, Veterinary Diagnostic
Laboratory, Manitoba Agriculture, 545 University Crescent,
Winnipeg, Manitoba R3T 5S6, Canada
FEATURES
source
1. .1768
/organism="porcine circovirus type 2-C"
/virus
/db_xref="taxon:85543"
/notes="sequence obtained from several overlapping PCRs
using DNA extracted from lung, mesenteric lymph node and
tonsil of pig; similar to Porcine circovirus sequence
presented in GenBank Accession Number AF027217; type-C
designation is based upon restriction endonuclease
digestion pattern"
13..35
repeat_region
/rpt_type=tandem
/rpt_unit=13..18
51..995
/notes="ORF-1."
/codon_start=1
/product="putative Rep and coat protein"
/protein_id="AAD03061.1"
/db_xref="PID:94106886"
/db_xref="GI:4106886"
/translation="MPSKNGRGQPHKRWFTLNNSSEDERKKIRELPISLFDYFI
VGEENEGRTPLHGFANFVKOTFNKVLGARCHIEKAGCTDOONKEYCSKEG
LLIQCAPRSQGRSLSLSTAVSLLESGLTVAEQHPVTFVFRNGLAELLKYSGRM
QKRDKNVIRVIGPPCGSKAANFADPTTTPKPRKNWMDYHGEEVVVDDYF
GWLPPDLDLCRDYPLTVETKGGPVFLARSILITNQTPLEWYSTAVPAVEALYR
RITSLVFNKNATEQSTEEGGQFVTLSPPCPEFFEYNY"
327..332
complement(357..671)
/codon_start=1
/product="ORF-3"
/protein_id="AAD03063.1"
/db_xref="PID:94106888"
/db_xref="GI:4106888"
/translation="MTVTPILVSRWFVPCGRVCKISSPFAFATPRPHNDVIRLPI
TLLHFAHFQKFSQPAEISDKRYRVLNCGHQTALQOGHSSQVTSLSRSSTL
TLLHFAHFQKFSQPAEISDKRYRVLNCGHQTALQOGHSSQVTSLSRSSTL

184 rpLeuArgLeuGlnThrSerGlyAsnValAspHisValGlyLeuGlyThr 200
/complement(386..565)
/codon_start=1
/product="ORF-4"
/protein_id="AAD03064.1"
/db_xref="PID:94106889"
/db_xref="GI:4106889"
/translation="MTCTFFVQSRFCIFPLTFKSSASPRKFLTNVTGCCSATVRLPL
SNKVLTAVDRLRCP"
553..732
/codon_start=1
/product="ORF-12"
/protein_id="AAD03070.1"
/db_xref="PID:94106895"
/db_xref="GI:4106895"
/translation="MTYSLMGHLGVAKANGLLILQTRKPHGNHLETSGGMVTVKKW
LLLMTFMVGCRGIY"
983..988
complement(989..1033)
/codon_start=1
/product="ORF-11"
/protein_id="AAD03069.1"
/db_xref="PID:94106894"
/db_xref="GI:4106894"
/translation="MNNKNHYEVIKKTQ"
1016..1177
/codon_start=1
/product="ORF-5"
/protein_id="AAD03065.1"
/db_xref="PID:94106890"
/db_xref="GI:4106890"
/translation="MVPIIHLGFWGVKIKFSELYIHGYTDIVLVVYTVFERSAE
IMVYISSSL"
complement(1022..1027)
complement(1256..1735)
/notes="similar to Porcine circovirus ORF-2 encoded by the
sequence presented in GenBank Accession Number AF027217."
/codon_start=1
/product="ORF-2"
/protein_id="AAD03062.1"
/db_xref="PID:94106887"
/db_xref="GI:4106887"
/translation="MTYPRRYRRRTSPRSHLGGILRRRPLVPRHYVRKNGI
FNTLSRTFGYTVATVTRPSWAVMLRFKIDDFVPGGTNKISIFFYRIRKVK
VEFWPCPITGGDRGVSSAVILDONFVPKATALTYDPIVYNSRHTITQPFYSHR"
complement(1519..1611)
/codon_start=1
/product="ORF-6"
/protein_id="AAD03066.1"
/db_xref="PID:94106891"
/db_xref="GI:4106891"
/translation="MASSTPASPAPSDILSRPQSERPPGRWTC"
1524..1631
/codon_start=1
/product="ORF-10"
/protein_id="AAD03068.1"
/db_xref="PID:94106893"
/db_xref="GI:4106893"
/translation="MSTAQEGVLTVALTVPKVRRLVKMPPELLQR"
1682..1741
/codon_start=1
/product="ORF-7"
/protein_id="AAD03067.1"
/db_xref="PID:94106892"
/db_xref="GI:4106892"
/translation="MAGAGPSSAVTPPWIRHS"
join(1750..1768,1..13)
/notes="putative replication site"
join(1762..1768,1..2)
/notes="similar to the nonnucleotide motif of Porcine
circovirus presented in GenBank Accession Numbers AF027217
and U49186"
449 a 364 c 490 g 465 t
BASE COUNT
```

ORIGIN

alignment\_scores: Quality: 1207.00 Length: 233  
Ratio: 5.341 Gaps: 0  
Percent Similarity: 96.996 Percent Identity: 94.850

alignment\_block:  
US-09-209-961-5 x AF109398/rev ..

Align seg 1/1 to reverse of: AF109398 from: 1 to: 1768

1 MetThrTyrProArgArgTyrArgArgArgHisArgProArgSe 17  
1735 ATGACGTATCCAAAGGAGGCTTACCGACAGAAGAGGACCGCCCGCAG 1686

17 rHisLeuGlyGlnIleLeuArgArgProTrpLeuValHisProArgH 34  
1685 CCATCTTGCCAGATCTCCCGCGCGCCCTGGCTCGTCCACCCCGCC 1636

34 isArgTyrArgTrpArgArgLysAsnGlyIlePheAsnThrArgLeuSer 50  
1635 ACCGTACCGCTGGAGAGGAATAATGGCATCTTCAACACCGCCCTCTCC 1586

51 ArgThrPheGlyTyrThrValLysArgThrThrValThrProSerTr 67  
1585 CGCACCTTCGGATATACTGTCAAGGCTACCCAGTCAAGACCGCTCCCTG 1536

67 palValAlaMetMetArgPheLysIleAspPheValProGlyG 84  
1535 GCGGTGGACATGCTGAGATTTAAATTCAGCACTTTGTTCGCCAGGAG 1486

84 lyGlyThrAsnLysIleSerIleProPheGluTyrTyrArgIleArgLys 100  
1485 GGGGACCAACAATCTATACCTTTGAATACTACAGATAAGAA 1436

101 ValLysValGluPheTrpProCysSerProIleThrGlnGlyAspArgG 117  
1435 GTTAAGTTGAATCTGGCCCTGCTCCCAATCACCAGGCTGACAGGG 1386

117 yValGlySerThrAlaValIleLeuAspAspAsnPheValThrLysAla 134  
1385 AGTTGGATCCAGTGTGTTATTCTAGATGATACTTTGTACCAAGGCCA 1336

134 hrAlaLeuThrTyrAspProTyrValAsnTyrSerArgHisThrIle 150  
1335 CAGCCCTAACCTATGACCCCTATGTAATACTCTCTCCGCCATACCA 1286

151 ProGlnProPheSerTyrHisSerArgTyrPheThrProLysProValLe 167  
1285 ACCCAGCCCTCTCTCACCACCTCCGTTAATTCACACCAACCTCTCT 1236

167 uAspSerThrIleAspTyrPheGlnProAsnAsnLysArgAsnGlnLeu 184  
1235 TGATGGACATCGATTCTTCAACCCCAATAACAAAGAAATCACTCT 1186

184 rPLeuArgLeuGlnThrSerGlyAsnValAspHisValGlyLeuGlyThr 200  
1185 GGCTGAGACTACAACCTACTGGAATAGTAGACCATGTAGCCCTCGCAT 1136

201 AlaPheGluAsnSerLysTyrAspGlnAspTyrAsnIleArgValThrMe 217  
1135 CGGTCGAAACAGATATATACCGCAGGACTACAATATCCGTGTAACAT 1086

217 tTyrValGlnPheArgGluPheAsnLeuLysAspProProLeuGluPro 233  
1085 GTATGTACAAATTCAGAGATTTAATCTTAAGACCCCGCCACTTAACCC 1037

seq\_name: gb\_vi:AF117753

seq\_documentation\_block:  
LOCUS AF117753 1768 bp DNA circular VRI 04-FEB-1999

```

/product="ORF-8"
/protein_id="AAD12312.1"
/db_xref="PID:g4219098"
/db_xref="GI:4219098"
/translation="MDTDHTVSVVDHPTAAASHKSHQ"
983..988
polyA_signal
CDS
complement(989..1033)
/codon_start=1
/product="ORF-11"
/protein_id="AAD12313.1"
/db_xref="PID:g4219099"
/db_xref="GI:4219099"
/translation="MNNKNHYEIKKTQ"
1016..1177
CDS
/codon_start=1
/product="ORF-5"
/protein_id="AAD12314.1"
/db_xref="PID:g4219100"
/db_xref="GI:4219100"
/translation="MVFIHLGFKWGVFKIKFSELYIHGYTDIVLVLVVYTVFERSAE"
YMWYISSSL"
complement(1022..1027)
polyA_signal
CDS
complement(1256..1735)
/note="similar to Porcine circovirus ORF-2 product encoded
by the sequence presented in GenBank Accession Number
AF027217"
/codon_start=1
/product="ORF-2"
/protein_id="AAD12315.1"
/db_xref="PID:g4219101"
/db_xref="GI:4219101"
/translation="WTFPRRRYRRRRSHLGOILRRPWLHRRHRYVWRKNGI
FNTRLSRTGYVKAIVTPSPWMDLRFKIDDFVPPGGTNKISIPPEYIRKVK
VFEPSPQITQDGRGVSSAVILDDNFFPKTALTALYVYVNSRHTITQPFYSYHR"
1524..1631
CDS
/codon_start=1
/product="ORF-10"
/protein_id="AAD12317.1"
/db_xref="PID:g4219103"
/db_xref="GI:4219103"
/translation="MSTAQEGVLTVVVALYVYKVRERRVLKMPFFLLQR"
complement(1528..1611)
CDS
/codon_start=1
/product="ORF-6"
/protein_id="AAD12316.1"
/db_xref="PID:g4219102"
/db_xref="GI:4219102"
/translation="MASSTPASPAPSILSLRLOPSPRPPGR"
1682..1741
CDS
/codon_start=1
/product="ORF-7"
/protein_id="AAD12318.1"
/db_xref="PID:g4219104"
/db_xref="GI:4219104"
/translation="MAAGAVSSAVTPPKRHS"
join(1750..1768,1..13)
stem_loop
/note="putative replication site"
join(1762..1768,1..2)
rep_origin
/note="similar to the nonanucleotide motif of Porcine
circovirus presented in GenBank Accession Numbers AF027217
and U49186."
452 a 358 c 497 g 461 t
BASE COUNT
ORIGIN

```

```

alignment_scores:
  Quality: 1202.00
  Ratio: 5.319
  Percent Similarity: 96.996
  Percent Identity: 93.562
alignment_block:
  US-09-209-961-5 x AF117753/rev ..

```

```

Align seg 1/1 to reverse of: AF117753 from: 1 to: 1768
1 MetThrTyrProArgArgArgTyrArgArgArgArgHisArgProArgSe 17
1735 ATGACGTTTCCAAAGGAGGGTTACCGCAGACAGACACCCCGCGCAG 1686
17 rHisLeuGlyGlnIleLeuArgArgProTyrPheValHisProArgH 34
1685 CCATTTGGCCAGATCCTCCGCGCGCCCTGGTTCTCTCCACCCCGCC 1636
34 IsArgTyrArgTyrArgArgLysAsnGlyIlePheAsnThrArgLeuSer 50
1635 ACCGTTACCGCTGGAGAGGAAAAATGGCATCTTCAACACCCGCTCTCC 1586
51 ArgThrPheGlyTyrThrValLysArgThrValThrProSerTr 67
1585 CGACCTTCGGATATCTGCAAGGCTACACAGTCAGAACGCCCTCTG 1536
67 pAlaValAspMetMetArgPheLysIleAspAspPheValProGlyG 84
1535 GCGGTAGACATGCTGAGATTAAATATAGACACTTTGTTCCCGGGAG 1486
84 lYgLyThrAsnLysIleSerIleProPheGlyTyrTyrArgIleArgLys 100
1485 GGGGACCAACAATCTCTATACCTTTGAATACTACAGATAAGAAAG 1436
101 ValLysValGluPheThrProCysSerProIleThrGlnGlyAspArgG 117
1435 GTTAAGGTTGAATCTGCGCCCTGCTCCCAATCACCCAGGTGACAGGG 1386
117 ValGlySerThrAlaValIleLeuAspAspAsnPheValThrLysAla 134
1385 AGTTGGATCCAGTGTGTTATTCTAGATGATAACTTTTCCCAAGAGCA 1336
134 hAlaLeuThrTyrAspProTyrValAsnTyrSerArgHisThrIle 150
1335 CAGCCCTAACCTATGACCCCTATGTAACCTACTCTCCCGCCATACCAT 1286
151 ProGlnProPheSerTyrHisSerArgTyrPheThrProLysProVal 167
1285 ACCCAACCCCTTTCTTACCACCTCCCGTTAAATTCACACCCCAACCTG 1236
167 uAspSerThrIleAspTyrPheGlnProAsnAsnLysArgAsnGlnLeu 184
1235 TGATTCGACATCGATTACTTCCAGCCCAATAACAAAGAAATCAACTCT 1186
184 rPLeuArgLeuGlnThrSerGlyAsnValAspHisValGlyLeuGlyThr 200
1185 GGCTGAGACTACAACCTACTGGAATGTAGACCATGTAGCCCTCGCACT 1136
201 AlaPheGluAsnSerLysTyrAspGlnAspTyrAsnIleArgValThrMe 217
1135 GCGTTCGAAACAGTATATACGACCCAGGACTACAATATCCGTGTAACCAT 1086
217 tHnValGlnPheArgGluPheAsnLeuLysAspProLeuGluPro 233
1085 GTATGTACAATTACAGAAATTAATCTTAAAGACCCCCCACTTAACCC 1037
seq_name: gb_v1:AF118097
seq_documentation_block:
LOCUS AF118097 700 bp DNA VRL 27-JUL-1999
DEFINITION Porcine circovirus type 2 isolate IAF-4370 unknown gene.
ACCESSION AF118097
NID 95596427
VERSION AF118097.1 GI:5596427
KEYWORDS porcine circovirus type 2.
SOURCE porcine circovirus type 2
ORGANISM Viruses; ssDNA viruses; Circoviridae; Circovirus.
REFERENCE 1 (bases 1 to 700)
AUTHORS Ouardani,M., Wilson,L., Jette,R., Montpetit,C. and Dea,S.
TITLE Multiplex PCR for detection and typing of porcine circoviruses

```

J. Clin. Microbiol. (1999) In press  
REFERENCE 2 (bases 1 to 700)  
AUTHORS Ouardani, M., Wilson, L., Jette, R., Montpetit, C. and Dea, S.  
TITLE Direct Submission  
JOURNAL Submitted (06-JAN-1999) CRMB, INRS-IAF, 531, Boulevard des  
Prairies, Laval, Que H7N 4Z3, Canada  
FEATURES  
source Location/Qualifiers  
1. 700  
/organism="porcine circovirus type 2"  
/isolate="IAF-4370"  
/db\_xref="taxon:85708"  
/note="virus isolated from lung, lymph node, spleen and  
tonsil tissue from pigs affected by postweaning  
multisystemic wasting syndrome"  
CDS complement(42..>700)  
/note="ORF2"  
/codon\_start=3  
/product="unknown"  
/protein\_id="AAD45581.1"  
/db\_xref="PID:g5596428"  
/db\_xref="GI:5596428"  
/translation="RSLGQILRRPWLHPHRYRWRKNGIENRSLRSTFGYTVKA  
TVTRPSWADMRPNIDDFVPPCGGYNKISIPFEYRIRKVKVEFWPCSPITOGDRG  
VGSTAVLDNFTKATLYDPVYNYSSRHTIPQFYSYHSRYETPKPVLDSITDIYFQ  
PNKRNQLMRLQTSRNVHVLGLTAFENSKYDQDYNIRVTMYVQFREFNLKDPPLKP  
"

BASE COUNT 156 a 127 c 210 g 207 t  
ORIGIN

alignment\_scores:  
Quality: 1163.00 Length: 218  
Ratio: 5.409 Gaps: 0  
Percent Similarity: 98.624 Percent Identity: 97.248

alignment\_block:  
US-09-209-961-5 x AF118097/rev ..

Align seg 1/1 to reverse of: AF118097 from: 1 to: 700

16 ArgSerHisLeuGlyGlnIleLeuArgArgProTrpLeuValHisPr 32  
|||||  
698 CGCAGCCATCTGGCAGATCTCCGCCGCCCGCTGGCTGCTCCACC 649  
|||||  
32 oArgHisArgTrpArgTrpArgGlyAsnGlyIlePheAsnThrArgL 49  
|||||  
648 CCGACACCGCTACCGTGGAGAGGAAATAATGGCATCTTCAACACCCGCC 599  
|||||  
49 euSerArgThrPheGlyTyrThrValLysArgThrValThrPro 65  
|||||  
598 TCTCCCGCACCTTCGGATATCTGCAAGGTCACACAGTACAGACGCC 549  
|||||  
66 SerTrpAlaValAspMetMetArgPheLysIleAspAspPheValPro 82  
|||||  
548 TCCTGGCGGTGGACATGAGATTAATATGACGACTTGTTCGCC 499  
|||||  
82 oGlyGlyThrAsnLysIleSerIleProPheGluTyrTyrArgIleA 99  
|||||  
498 GGGAGGGGGGACCAAAATCTATATACCTTTGAATACTACAGATAA 449  
|||||  
99 rgLysValLysValGluPheTrpProCysSerProIleThrGlnGlyAsp 115  
|||||  
448 GAAAGGTTAAGGTTGAATTCCTGCCCTGCTCCGCCATCACCAGGCTGAC 399  
|||||  
116 ArgGlyValGlySerThrAlaValIleLeuAspAspAsnPheValThrly 132  
|||||  
398 AGGGAGTGGGCTCCACGCTGCTGTTATCTAGATGATACTTTGTACAAA 349  
|||||  
132 sAlaThrAlaLeuThrTyrAspProTyrValAsnTyrSerSerArgHis 149  
|||||  
348 GGCCACAGCCCTAACCTATGACCCATATGTAACACTACTCTCCCGGCATA 299  
|||||  
149 hrIleProGlnProPheSerTyrHisSerArgTyrPheThrProLysPro 165

|||||  
298 CAATCCCCCAACCTTCTCTACCACTCCCGTTACTTCACACCCAAACCT 249  
|||||  
166 ValLeuAspSerThrIleAspTyrPheGlnProAsnAsnLysArgAsnG 182  
|||||  
248 GTCTTGGACTCCACTATTGATTCTTCCAAACCAATACAAAGGAATCA 199  
|||||  
182 nLeuTrpLeuArgLeuGlnThrSerGlyAsnValAspHisValGlyLeu 199  
|||||  
198 GCITTTGGATGAGGCTACAAACCTCTAGAAATGTGACCACTAGGCCTCG 149  
|||||  
199 lyThrAlaPheGluAsnSerLysTyrAspGlnAspTyrAsnIleArgVal 215  
|||||  
148 GCACCTGGCTTCGAAAACAGTAAATACGACGAGGACTACAATATCCGTGA 99  
|||||  
216 ThrMetTyrValGlnPheArgGluPheAsnLeuLysAspProLeuG 232  
|||||  
98 ACCATGTATGTACAATTCAGAGAATTTAATCTTAAGACCCCCCCTTAA 49  
|||||  
232 upro 233  
48 ACCC 45  
:|||

---



OM of: US-09-209-961-5 to: N\_Geneseq\_36:\* out\_format : pfs

Date: Dec 28, 1999 1:50 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-1998 CompuGen Ltd.

Command line parameters:

-MODEL=framet.p2n.model -DEV=xlp  
-Q/cgn2\_1/USPTO\_spool/US0209961/runat\_22121999\_101043\_17942/app\_query.fasta.1  
-DB=N\_Geneseq\_36 -QMT=fastap -SUFFIX=ring -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -QGAPEXT=0.500 -XGAPOP=10.000 -XGAPEXT=0.500  
-QGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blossum62  
-TRANS-human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=escore  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09209961  
-NCPU=6 -ICPU=3 -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-209-961-5

Query length: 233

Database: N\_Geneseq\_36:\*

Database sequences: 311585

Database length: 125096042

Search time (sec): 140.030000

score_list:	Sequence	Strd	Orig	zScore	Escore	Len	Documentation
N_Geneseq_36:X35213	-	692.00	1297.36	1.2e-64	1768	1	Nucleotide sequence of PCV iso
N_Geneseq_36:X35212	-	681.00	1276.30	1.7e-63	1768	1	Genomic DNA sequence of PCV iso
N_Geneseq_36:X35211	-	681.00	1276.30	1.7e-63	1768	1	Nucleotide sequence of PCV iso
N_Geneseq_36:X35210	-	669.00	1253.33	3.3e-62	1767	1	Nucleotide sequence of PCV iso
N_Geneseq_36:X35211	-	669.00	1253.33	3.3e-62	1767	1	Nucleotide sequence of PCV iso
N_Geneseq_36:X35211	-	478.00	887.73	7.6e-42	1759	1	Nucleotide sequence which has
N_Geneseq_36:X35211	-	91.50	147.30	1.33	1848	1	A. niger glucose-oxidase gene.
N_Geneseq_36:X35211	-	91.50	147.30	1.33	1848	1	Aspergillus niger glucose oxid
N_Geneseq_36:X35211	-	91.50	146.34	1.50	2024	1	Nucleotide sequence of cDNA fr
N_Geneseq_36:X35211	-	87.50	145.97	1.57	1015	1	PEBP2 alpha A gene expression
N_Geneseq_36:X35211	-	90.00	145.08	1.76	1738	1	Genomic sequence encoding huma
N_Geneseq_36:X35211	-	83.50	141.90	2.65	7222	1	Polynucleotide sequence from th
N_Geneseq_36:X35211	-	91.50	139.91	3.42	3722	1	Non-B, non-C, non-G hepatitis
N_Geneseq_36:X35211	-	87.00	139.33	3.68	1738	1	Alpha 1b adrenergic receptor g
N_Geneseq_36:X35211	-	87.00	139.33	3.68	1738	1	Alpha-1b adrenergic receptor g
N_Geneseq_36:X35211	-	83.50	136.15	5.54	1245	1	Treponema pallidum (TP) 47 kDa
N_Geneseq_36:X35211	-	83.50	135.81	5.79	1286	1	Coding sequence for modified T
N_Geneseq_36:X35211	-	86.50	135.46	6.05	2291	1	Entire amylase gene. Expressi
N_Geneseq_36:X35211	-	83.50	135.44	6.07	1332	1	Wild type T. pallidum 47K anti
N_Geneseq_36:X35211	-	83.50	135.16	6.29	1368	1	Nucleic acid-bound polypeptide
N_Geneseq_36:X35211	-	83.50	135.09	6.35	1377	1	Treponema pallidum 47 kDa surf
N_Geneseq_36:X35211	-	83.50	135.09	6.35	1377	1	Treponema pallidum 47 kDa anti
N_Geneseq_36:X35211	-	76.50	134.55	6.80	407	1	Mycobacterium tuberculosis anti
N_Geneseq_36:X35211	-	76.50	134.55	6.80	407	1	M. tuberculosis immunogenic pol
N_Geneseq_36:X35211	-	77.50	134.42	6.92	494	1	Mycobacterium tuberculosis anti
N_Geneseq_36:X35211	-	77.50	134.42	6.92	494	1	M. tuberculosis immunogenic pol
N_Geneseq_36:X35211	-	87.00	134.11	7.20	2852	1	Mycobacterium tuberculosis ant
N_Geneseq_36:X35211	-	87.00	134.11	7.20	2852	1	M. tuberculosis immunogenic po
N_Geneseq_36:X35211	-	77.50	133.58	7.71	535	1	Mycobacterium tuberculosis anti
N_Geneseq_36:X35211	-	77.50	133.58	7.71	535	1	M. tuberculosis immunogenic pol
N_Geneseq_36:X35211	-	83.50	133.09	8.20	1664	1	Coding sequence for modified T
N_Geneseq_36:X35211	-	92.50	130.97	10.76	10411	1	Streptococcus pneumoniae genc
N_Geneseq_36:X35211	-	80.50	129.69	12.69	1333	1	Mouse neurogenic differentiat
N_Geneseq_36:X35211	-	80.00	128.26	13.41	1268	1	Human neurogenic differentiat
N_Geneseq_36:X35211	-	82.50	129.26	13.41	1268	1	DNA encoding human neuroD3 pro
N_Geneseq_36:X35211	-	82.50	127.90	15.97	2271	1	Amylase gene from Streptomyces
N_Geneseq_36:X35211	-	79.00	127.45	16.92	1256	1	P. velutina lectin cDNA. Sandw
N_Geneseq_36:X35211	-	85.00	127.42	16.97	3739	1	Non-B, non-C, non-G hepatitis
N_Geneseq_36:X35211	-	79.50	126.95	18.02	1441	1	M. tuberculosis antigen clone
N_Geneseq_36:X35211	-	79.50	126.95	18.02	1441	1	M. tuberculosis antigen clone
N_Geneseq_36:X35211	-	81.00	126.78	18.43	1923	1	Mycobacterium tuberculosis and

N\_Geneseq\_36:V64551 + 81.00 126.78 18.43 1923 ! M. tuberculosis immunogenic  
N\_Geneseq\_36:Q12567 + 79.50 126.44 19.25 1513 ! Alkaline protease cDNA from  
N\_Geneseq\_36:Q14478 + 83.50 125.35 22.13 3465 ! HSVgB gene. HSVgB polypepti

seq\_name: N\_Geneseq\_36:X35213

seq\_documentation\_block:

ID X35213 standard; DNA; 1768 BP.

AC X35213;

DT 01-JUL-1999 (first entry)

DE Nucleotide sequence of PCV isolate Impl010.

KW PCV isolate; type II porcine circovirus; PCV; PMWS;

OS Porcine multisystemic wasting syndrome; pig; vaccine; ss.

PN FR2769322-Al.

PD 09-APR-1999.

PF 22-JAN-1998; 000873.

PR 03-OCT-1997; FR-012382.

PA (MERI-) MERIAL SAS.

PI Allan G. Chapuis GE, Charreyre CE, Clark E, Ellis J,

PI Haines D, Harding J, Hassard L, Meehan B;

DR WPI; 99-246948/21.

PT New type II porcine circovirus, used for, e.g. passive immunization

PS Claim 14; Fig 4; 48pp; French.

CC The present sequence represents the nucleotide sequence of PCV isolate

CC Impl010. The specification describes a preparation of type II

CC porcine circovirus (PCV), which is particularly isolated from a lesion,

CC from a pig with symptoms of PMWS (porcine multisystemic wasting

CC syndrome). PCV (attenuated or inactivated), polypeptides derived from

CC it, and vectors that express these polypeptides are all useful in

CC vaccines, suitable for administration to adult or young pigs, or to

CC pregnant sows (for passive immunization of their offspring). DNA

CC isolated from PCV is used for in vivo or in vitro expression of viral

CC polypeptides, also as probes or primers for diagnosis in usual

CC hybridization or amplification assays. These polypeptides may also be

CC used diagnostically to detect PCV-specific antibodies, while antibodies

CC raised against the polypeptides can be used to detect antigens, in any

CC usual immunoassay format.

CC Sequence 1768 BP; 451 A; 361 C; 495 G; 461 T;

alignment\_scores:

Quality: 592.00 Length: 128

Ratio: 5.406 Gaps: 0

Percent Similarity: 100.00 Percent Identity: 98.438

alignment\_block:

US-09-209-961-5 x X35213/rev ..

Align seg 1/1 to reverse of: X35213 from: 1 to: 1768

106 TtrProCysSerProIleThrGlnGlyAspArgGlyValGlySerThrAl 122

|||||

1767 TGGCCCTGCTCCCCCATCCACCCAGGGGTAGGGGAGTGGGCTCCACTGC 1718

122 aValIleLeuAspAspAsnPhelValThrLysAlaThrAlaLeuThrTyra 139

|||||

1717 TGTATTCTAGATGATAACTTTGTAACAAGGCCACAGCCCTAACCTATG 1668

139 spProTyValAsnTySerSerArgHisThrIleProGlnProPheser 155

|||||

1667 ACCCATGTAAACTACTCTCCGCCATACATAATCCGCCAACCTCTTCC 1618

156 TyHisSerArgTyPheThrProLysProValLeuAspSerThrIleAs 172

|||||

1617 TACCACCTCCCGTACTTACACACCAACCTGTTCTTGACTCCACTATGA 1568

172 pTyHisGlnProAsnAsnLysArgAsnGlnLeuTyPleuArgLeuGlnT 189

|||||

1567 TTACTTCCACCAATAACAAGGATCAGCTTGGCTGAGACTACAAA 1518

189 hrSerGlyAsnValAspHisValGlyLeuGlyThrAlaPheGluAsnSer 205

```

1517 CCTCTGGAATGTGACACGCTAGCGCTCGGCTCGCTGCAAAACAGT 1468
|||||
206 LysTyrAspGlnAspTyrAsnIleArgValThrMetTyrValGlnPheAr 222
|||||
1467 AAATACGACACAGGACTACAATATCGGTGAACCATGTATGTACAATTGAG 1418
|||||
222 gGluPheAsnLeuLysAspProLeuGluPro 233
|||||
1417 AGAATTTAATCTTAAGACCCCTTAAACCC 1384
|||||

```

seq\_name: N\_Geneseq\_36:X35012

seq\_documentation\_block:

```

ID X35012 standard; DNA; 1768 BP.
AC X35012;
DT 01-JUL-1999 (first entry)
DE Genomic DNA sequence of PCV strain 999PCV.
KW PCV; strain PCV999; pig; PMWS; porcine multisystemic wasting syndrome;
KW vaccine; ss.
OS Porcine circovirus
PN FR2769321-AL.
PD 09-APR-1999.
PF 03-OCT-1997; 012382.
PR 03-OCT-1997; FR-012382.
PA (MERI-) MERIAL SAS.
PA (UYBE-) UNIV QUEENS BELFAST.
PA (UYSA-) UNIV SASKATCHEWAN.
PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,
PI Haines D, Harding J, Hassard L, Meehan B;
DR WPI: 99-246947/21.
PT New porcine circovirus from animals with porcine systemic wasting
PT syndrome
PS Claim 13; Fig 1; 35pp; French.
CC The present sequence represents the genomic sequence of porcine
CC circovirus (PCV) strain PCV999. The sequence is obtained from PCV
CC isolated from a physiological or tissue sample, particularly from
CC a lesion, from a pig showing symptoms of PMWS (porcine multisystemic
CC wasting syndrome), or cultured cells, infected with PCV isolated from
CC such pigs. PCV (attenuated or inactivated), polypeptides derived from
CC it, and vectors that express these polypeptides are all useful in
CC vaccines, suitable for administration to adult or young pigs, or to
CC pregnant sows (for passive immunization of their offspring). DNA
CC isolated from PCV is used for in vivo or in vitro expression of viral
CC polypeptides, also as probes or primers for diagnosis in usual
CC hybridization or amplification assays. These polypeptides may also be
CC used diagnostically to detect PCV-specific antibodies, while antibodies
CC raised against the polypeptides can be used to detect antigens, in any
CC usual immunoassay format.
SQ Sequence 1768 BP; 453 A; 359 C; 486 G; 455 T;

```

alignment\_scores:

```

Quality: 681.00 Length: 128
Ratio: 5.405 Gaps: 0
Percent Similarity: 98.438 Percent Identity: 97.656

```

alignment\_block:

US-09-209-961-5 x X35012/rev ..

Align seg 1/1 to reverse of: X35012 from: 1 to: 1768

```

106 TrpProCysSerProIleThrGlnGlyAspArgGlyValGlySerThrAl 122
|||||
1768 TGGCCCTCTCTCCCATCCAGGGGTAGGGGAGTGGGCTCCACTGC 1719
|||||
122 aValIleLeuAspAspAsnPheValThrLysAlaThrAlaLeuThrTyra 139
|||||
1718 TGTATTCTAGATGATACTTTGTACAAAGGCCACAGCCCTAAACCTATG 1669
|||||
139 sPProTyrValAsnTyrSerSerArgHisThrIleProGlnProPheSer 155
|||||
1668 rECCCATATGTAACACTACTCTCTCCCGCATACATATCCCAACCCCTCTCC 1619
|||||

```

```

156 TyrHisSerArgTyrPheThrProLysProValLeuAspSerThrIleAs 172
|||||
1618 TACCACTCCCGTTACTTCACACCCAAACCTGTTCTTGACTCCCACTATGA 1569
|||||
172 pTyrPheGlnProAsnAsnLysArgAsnGlnLeuTrpLeuArgLeuGlnT 189
|||||
1568 TTACTTCCCAACCAATAACAAAGGAATCAGCTTGTGCTGAGGCTACAAA 1519
|||||
189 hrSerGlyAsnValAspHisValGlyLeuGlyThrAlaPheGluAsnSer 205
|||||
1518 CCTCTAGAAATGTGGACACGCTAGGCTCGGCACCTGCGTTTCGAAACAGT 1469
|||||
206 LysTyrAspGlnAspTyrAsnIleArgValThrMetTyrValGlnPheAr 222
|||||
1468 ATATACGACACGAGCTACAATATCCGTGTAAACCATGTATGTACAATTGAG 1419
|||||
222 gGluPheAsnLeuLysAspProLeuGluPro 233
|||||
1418 AGAATTTAATCTTAAGACCCCTTAAACCC 1385
|||||

```

seq\_name: N\_Geneseq\_36:X35212

seq\_documentation\_block:

```

ID X35212 standard; DNA; 1768 BP.
AC X35212;
DT 01-JUL-1999 (first entry)
DE Nucleotide sequence of PCV isolate Imp999 (corrected version).
KW PCV isolate; type II porcine circovirus; PCV; PMWS;
KW porcine multisystemic wasting syndrome; pig; vaccine; ss.
OS Porcine circovirus.
PN FR2769322-AL.
PD 09-APR-1999.
PF 22-JAN-1998; 000873.
PR 03-OCT-1997; FR-012382.
PA (MERI-) MERIAL SAS.
PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,
PI Haines D, Harding J, Hassard L, Meehan B;
DR WPI: 99-246948/21.
PT New type II porcine circovirus, used for, e.g. passive immunization
PT of pregnant sows
PS Claim 14; Fig 3; 48pp; French.
CC The present sequence represents the nucleotide sequence of PCV isolate
CC Imp999. The specification describes a preparation of type II
CC porcine circovirus (PCV), which is particularly isolated from a lesion,
CC from a pig with symptoms of PMWS (porcine multisystemic wasting
CC syndrome). PCV (attenuated or inactivated), polypeptides derived from
CC it, and vectors that express these polypeptides are all useful in
CC vaccines, suitable for administration to adult or young pigs, or to
CC pregnant sows (for passive immunization of their offspring). DNA
CC isolated from PCV is used for in vivo or in vitro expression of viral
CC polypeptides, also as probes or primers for diagnosis in usual
CC hybridization or amplification assays. These polypeptides may also be
CC used diagnostically to detect PCV-specific antibodies, while antibodies
CC raised against the polypeptides can be used to detect antigens, in any
CC usual immunoassay format.
SQ Sequence 1768 BP; 453 A; 360 C; 494 G; 461 T;

```

alignment\_scores:

```

Quality: 681.00 Length: 128
Ratio: 5.405 Gaps: 0
Percent Similarity: 98.438 Percent Identity: 97.656

```

alignment\_block:

US-09-209-961-5 x X35212/rev ..

Align seg 1/1 to reverse of: X35212 from: 1 to: 1768

```

106 TrpProCysSerProIleThrGlnGlyAspArgGlyValGlySerThrAl 122
|||||
1767 TGGCCCTCTCTCCCATCCAGGGGTAGGGGAGTGGGCTCCACTGC 1718
|||||

```

122 aValileLeuAspAsnPhelThrLysAlaThrLysAlaLeuThrTyrA 139  
 |||||  
 1717 TGTATTCTAGATGATACTTTGTACAAAGGCCACAGCCCTAACCTATG 1668  
 139 spProTyrValAsnTyrSerArgHisThrIleProGlnProPheSer 155  
 |||||  
 1667 ACCCATATGTAAACTACTCTCCCGCATACATCCCAACCCCTCTCC 1618  
 156 TyrHisSerArgTyrPheThrProLysProValLeuAspSerThrIleAs 172  
 |||||  
 1617 TACCACCTCCCGTTACTTACACCCCAACCTGTCTTGACTCCACTATTGA 1568  
 172 pTyrPheGlnProAsnAsnLysArgAsnGlnLeuTyrPheArgLeuGlnT 189  
 |||||  
 1567 TTACTTCCACCAACAATAACAAGGAATCAGCTTTGGCTGAGGCTACAAA 1518  
 189 hrSerGlyAsnValAspHisValGlyLeuGlyThrAlaPheGluAsnSer 205  
 |||||  
 1517 CCTCTAGAAATGTGGACCACTAGGCTCGGCACTCGCTTCGAACACAGT 1468  
 206 LysTyrAspGlnAspTyrAsnIleArgValThrMetTyrValGlnPheAr 222  
 |||||  
 1467 ATATACGACGAGGACTACAAATATCCGTGTAAACCATGTATGTACAAATTCAG 1418  
 222 gGluPheAsnLeuLysAspProProLeuGluPro 233  
 |||||  
 1417 AGAATTTAATCTTAAGACCCCACTTAACCC 1384

seq\_name: N\_Geneseq\_36:X35214

seq\_documentation\_block:

ID X35214 standard; DNA; 1768 BP.

AC X35214;

DT 01-JUL-1999 (first entry)

DE Nucleotide sequence of PCV isolate Imp999 (uncorrected version).

KW PCV isolate; type II porcine circovirus; PCV; PMWS;

KW porcine multisystemic wasting syndrome; pig; vaccine; ss.

OS Porcine circovirus.

PN FR2769322-A1.

PD 09-APR-1999.

PF 22-JAN-1998; 000873.

PR 03-OCT-1997; FR-012382.

PA (MERI-) MERIAL SAS.

PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,

PI Haines D, Harding J, Hassard L, Meehan B;

DR WPI: 99-246948/21.

PT New type II porcine circovirus, used for, e.g. passive immunization

PT of pregnant sows

PS Claim 14; Fig 6; 48pp; French.

CC The present sequence represents the nucleotide sequence of PCV isolate  
 CC Imp999. The specification describes a preparation of type II  
 CC porcine circovirus (PCV), which is particularly isolated from a lesion,  
 CC from a pig with symptoms of PMWS (porcine multisystemic wasting  
 CC syndrome). PCV (attenuated or inactivated), polypeptides derived from  
 CC it, and vectors that express these polypeptides are all useful in  
 CC vaccines, suitable for administration to adult or young pigs, or to  
 CC pregnant sows (for passive immunization of their offspring). DNA  
 CC isolated from PCV is used for in vivo or in vitro expression of viral  
 CC polypeptides, also as probes or primers for diagnosis in usual  
 CC hybridization or amplification assays. These polypeptides may also be  
 CC used diagnostically to detect PCV-specific antibodies, while antibodies  
 CC raised against the polypeptides can be used to detect antigens, in any  
 CC usual immunoassay format.

SQ Sequence 1768 BP; 453 A; 359 C; 486 G; 455 T;

alignment\_scores:

Quality: 681.00

Ratio: 5.405

Percent Similarity: 98.438

Percent Identity: 97.656

alignment\_block:

US-09-209-961-5 x X35214/rev ..

Align seg 1/1 to reverse of: X35214 from: 1 to: 1768

106 TrpProCysSerProIleThrGlnGlyAspArgGlyValGlySerThrAl 122  
 |||||  
 1768 TGGCCCTGCTCCCCCATCACCCAGGTGATAGGGAGTGGGCTCCACTGC 1719  
 122 aValileLeuAspAsnPhelThrLysAlaThrLysAlaLeuThrTyrA 139  
 |||||  
 1718 TGTATTCTAGATGATACTTTGTACAAAGGCCACAGCCCTAACCTATG 1669  
 139 spProTyrValAsnTyrSerArgHisThrIleProGlnProPheSer 155  
 |||||  
 1668 ACCCATATGTAAACTACTCTCCCGCATACATCCCAACCCCTCTCC 1619  
 156 TyrHisSerArgTyrPheThrProLysProValLeuAspSerThrIleAs 172  
 |||||  
 1618 TACCACCTCCCGTTACTTACACCCCAACCTGTCTTGACTCCACTATTGA 1569  
 172 pTyrPheGlnProAsnAsnLysArgAsnGlnLeuTyrPheArgLeuGlnT 189  
 |||||  
 1568 TTACTTCCACCAACAATAACAAGGAATCAGCTTTGGCTGAGGCTACAAA 1519  
 189 hrSerGlyAsnValAspHisValGlyLeuGlyThrAlaPheGluAsnSer 205  
 |||||  
 1518 CCTCTAGAAATGTGGACCACTAGGCTCGGCACTCGCTTCGAACACAGT 1469  
 206 LysTyrAspGlnAspTyrAsnIleArgValThrMetTyrValGlnPheAr 222  
 |||||  
 1468 ATATACGACGAGGACTACAAATATCCGTGTAAACCATGTATGTACAAATTCAG 1419  
 222 gGluPheAsnLeuLysAspProProLeuGluPro 233  
 |||||  
 1418 AGAATTTAATCTTAAGACCCCACTTAACCC 1385

seq\_name: N\_Geneseq\_36:X35210

seq\_documentation\_block:

ID X35210 standard; DNA; 1767 BP.

AC X35210;

DT 01-JUL-1999 (first entry)

DE Nucleotide sequence of PCV isolate Imp1011-48121.

KW PCV isolate; type II porcine circovirus; PCV; PMWS;

KW porcine multisystemic wasting syndrome; pig; vaccine; ss.

OS Porcine circovirus.

PN FR2769322-A1.

PD 09-APR-1999.

PF 22-JAN-1998; 000873.

PR 03-OCT-1997; FR-012382.

PA (MERI-) MERIAL SAS.

PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,

PI Haines D, Harding J, Hassard L, Meehan B;

DR WPI: 99-246948/21.

PT New type II porcine circovirus, used for, e.g. passive immunization

PT of pregnant sows

PS Claim 14; Fig 1; 48pp; French.

CC The present sequence represents the nucleotide sequence of PCV isolate  
 CC Imp1011-48121. The specification describes a preparation of type II  
 CC porcine circovirus (PCV), which is particularly isolated from a lesion,  
 CC from a pig with symptoms of PMWS (porcine multisystemic wasting  
 CC syndrome). PCV (attenuated or inactivated), polypeptides derived from  
 CC it, and vectors that express these polypeptides are all useful in  
 CC vaccines, suitable for administration to adult or young pigs, or to  
 CC pregnant sows (for passive immunization of their offspring). DNA  
 CC isolated from PCV is used for in vivo or in vitro expression of viral  
 CC polypeptides, also as probes or primers for diagnosis in usual  
 CC hybridization or amplification assays. These polypeptides may also be  
 CC used diagnostically to detect PCV-specific antibodies, while antibodies  
 CC raised against the polypeptides can be used to detect antigens, in any  
 CC usual immunoassay format.

SQ Sequence 1767 BP; 447 A; 360 C; 502 G; 458 T;

```

alignment_scores:
  Quality: 669.00      Length: 128
  Ratio: 5.310        Gaps: 0
  Percent Similarity: 98.438  Percent Identity: 95.312

alignment_block:
US-09-209-961-5 x X35210/rev ..

Align seg 1/1 to reverse of: X35210 from: 1 to: 1767

106 TrpProCysSerProIleThrGlnGlyAspArgGlyValGlySerThrAl 122
|||||
1766 TGGCCCTGCTCCCGATCACCAGGTCACAGGGAGTGGGCTCCAGTGC 1717
|||||
122 aValIleLeuAspAspAsnPhValThrLysAlaThrAlaLeuThrTyra 139
|||||
1716 TGTATTCTAGATGATACTTTGTAAACAAAGGCCACAGCCCTCACCTATG 1667
|||||
139 spProTyrValAsnTyrSerArgHisThrIleProGlnProPheSer 155
|||||
1666 ACCCTATGTAACTACTCTCCCGCATACCATACCCAGCCCTCTCTCC 1617
|||||
156 TyrHisSerArgTyrPheThrProLysProValLeuAspSerThrIleAs 172
|||||
1616 TACCACCTCCGCTACTTTACCCCAAAACCTGCTCTAGATTCCACTATTGA 1567
|||||
172 pTyrPheGlnProAsnAsnLysArgAsnGlnLeuTrpLeuArgLeuGlnT 189
|||||
1566 TTACTTCCAAACCAACAAACAAAGAAACAGCTGTGGCTGAGACTACAAA 1517
|||||
189 hrSerGlyAsnValAspHisValGlyLeuGlyThrAlaPheGluAsnSer 205
|||||
1516 CTACTGGAATGTAGACACGCTAGGCTCGGCCTCGCTAGATTCCACTATG 1467
|||||
206 LysTyrAspGlnAspTyrAsnIleArgValThrMetTyrValGlnPheAr 222
|||||
1466 ATATACGACGAGGAATACATATCCGTGAACCATGTATGTACATTTCAG 1417
|||||
222 gGluPheAsnLeuLysAspProProLeuGluPro 233
|||||
1416 AGAATTTAATCTTAAAGACCCCCCACTTAACCCCT 1383

seq_name: N_Geneseq_36.X35211

seq_documentation_block:
ID X35211 standard; DNA; 1767 BP.
AC X35211;
DT 01-JUL-1999 (first entry)
DE Nucleotide sequence of PCV isolate Imp1011-48285.
KW PCV isolate; type II porcine circovirus; PCV; PMWS;
KW porcine multisystemic wasting syndrome; pig; vaccine; ss.
OS Porcine circovirus.
PN FR2769322-Al.
PD 09-APR-1999.
PE 22-JAN-1998; 000873.
PR 03-OCT-1997; FR-012382.
PA (MERI-) MERIAL SAS.
PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,
PI Haines D, Harding J, Hassard L, Meehan B;
DR WPI: 99-246948/21.
PT New type II porcine circovirus, used for, e.g. passive immunization
PT of pregnant sows
PS Claim 14; Fig 2; 48pp; French.
CC The present sequence represents the nucleotide sequence of PCV isolate
CC Imp1011-48285. The specification describes a preparation of type II
CC porcine circovirus (PCV), which is particularly isolated from a lesion,
CC from a pig with symptoms of PMWS (porcine multisystemic wasting
CC syndrome). PCV (attenuated or inactivated), polypeptides derived from
CC it, and vectors that express these polypeptides are all useful in
CC vaccines, suitable for administration to adult or young pigs, or to
CC pregnant sows (for passive immunization of their offspring). DNA
CC isolated from PCV is used for in vivo or in vitro expression of viral
CC polypeptides, also as probes or primers for diagnosis in usual

```

```

alignment_scores:
  Quality: 669.00      Length: 128
  Ratio: 5.310        Gaps: 0
  Percent Similarity: 98.438  Percent Identity: 95.312

alignment_block:
US-09-209-961-5 x X35211/rev ..

Align seg 1/1 to reverse of: X35211 from: 1 to: 1767

106 TrpProCysSerProIleThrGlnGlyAspArgGlyValGlySerThrAl 122
|||||
1766 TGGCCCTGCTCCCGATCACCAGGTCACAGGGAGTGGGCTCCAGTGC 1717
|||||
122 aValIleLeuAspAspAsnPhValThrLysAlaThrAlaLeuThrTyra 139
|||||
1716 TGTATTCTAGATGATACTTTGTAAACAAAGGCCACAGCCCTCACCTATG 1667
|||||
139 spProTyrValAsnTyrSerArgHisThrIleProGlnProPheSer 155
|||||
1666 ACCCTATGTAACTACTCTCCCGCATACCATACCCAGCCCTCTCTCC 1617
|||||
156 TyrHisSerArgTyrPheThrProLysProValLeuAspSerThrIleAs 172
|||||
1616 TACCACCTCCGCTACTTTACCCCAAAACCTGCTCTAGATTCCACTATTGA 1567
|||||
172 pTyrPheGlnProAsnAsnLysArgAsnGlnLeuTrpLeuArgLeuGlnT 189
|||||
1566 TTACTTCCAAACCAACAAACAAAGAAACAGCTGTGGCTGAGACTACAAA 1517
|||||
189 hrSerGlyAsnValAspHisValGlyLeuGlyThrAlaPheGluAsnSer 205
|||||
1516 CTACTGGAATGTAGACACGCTAGGCTCGGCCTCGCTAGATTCCACTATG 1467
|||||
206 LysTyrAspGlnAspTyrAsnIleArgValThrMetTyrValGlnPheAr 222
|||||
1466 ATATACGACGAGGAATACATATCCGTGAACCATGTATGTACATTTCAG 1417
|||||
222 gGluPheAsnLeuLysAspProProLeuGluPro 233
|||||
1416 AGAATTTAATCTTAAAGACCCCCCACTTAACCCCT 1383

seq_name: N_Geneseq_36.X35013

seq_documentation_block:
ID X35013 standard; DNA; 1759 BP.
AC X35013;
DT 01-JUL-1999 (first entry)
DE Nucleotide sequence which has homology to PCV sequence.
KW PCV; strain PCV999; pig; PMWS; porcine multisystemic wasting syndrome;
KW vaccine; ss.
OS Sus sp.
PN FR2769321-Al.
PD 09-APR-1999.
PE 03-OCT-1997; 012382.
PR 03-OCT-1997; FR-012382.
PA (MERI-) MERIAL SAS.
PA (UYBE-) UNIV QUEENS BELFAST.
PA (UYSA-) UNIV SASKATCHEWAN.
PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,
PI Haines D, Harding J, Hassard L, Meehan B;
DR WPI: 99-246947/21.
PT New porcine circovirus from animals with porcine systemic wasting
PT syndrome
PS Disclosure; Fig 2; 35pp; French.
CC The specification describes a genomic sequence of porcine

```



```

139 pProTyrValAsnTyrSerSerArgHisThrIleProGlnProPheSert 156
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1407 CCTACTCTCACCAC.....TTGGCCT 1429

156 yHisSerArgTyrPheThrProLysProValLeuAspSerThrIleAsp 172
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1430 ACGACCCCTCAGTACTCTC..... 1449

173 TyrPheGlnProAsnAsnLysArgAsnGlnLeuTrpLeu..... 185
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1450 .....AACGAGCTGGACCTGCTCGTCAAGC 1475

186 ....ArgLeuGlnThrSerGlyAsnValAspHisValGly...LeuGlyT 200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1476 TGGCGTACTCAACTGCGCGCCCAACATCTCCAACCTCGGTGCCATGACAGA 1525

200 hrAlaPhe.....GluAsnSerLysTyrAspGln 209
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1526 CCTACTTCGCTGGGAGACTATCCCGGTGATACCTCGCGTATGATGCC 1575

210 AspTyrAsnIleArgValThrMetTyrValGlnPhe 221
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1576 GATTGTAGC...GCTGGACTGAGTACATCCCGTAC 1608

```

seq\_name: N\_Geneseq\_36:T29485

seq\_documentation\_block:

```

ID T29485 standard; DNA; 1848 BP.
AC T29485;
DE 20-NOV-1996 (first entry)
DE Aspergillus niger glucose oxidase gene.
KW Glucose oxidase; expression vector; transgenic plant; resistance;
KW fungal; bacterial; pathogen; potato; ds.
OS Aspergillus niger.
FH Key
FT Location/Qualifiers
FT 16..1833
FT cds
FT /*tag= a
PN US5516671-A.
PD 14-MAY-1996.
PF 24-NOV-1993; 161041.
PR 24-NOV-1993; US-161041.
PR 03-NOV-1994; US-333802.
PA (MONS.) MONSANTO CO.
PI Lawrence EB, Levine EB, Shah DM;
DR WPI; 96-251010/25.
DR P-PSDB; R95173.
PT DNA mol., comprising promoter, Aspergillus glucose oxidase gene and
PT polyA signal - used in prodn. of transgenic plants, esp. potato(s),
PT resistant to bacterial and fungal pathogens
PS Claim 2; Columns 13-18; 12pp; English.
CC The present sequence, which encodes A. niger glucose oxidase (AGO),
CC was used in the construction of a recombinant double stranded DNA
CC mol., comprising a promoter functional in plants, the present
CC sequence and a 3'-untranslated region, functional in plant cells,
CC which adds polyA nucleotides to the 3'-end of the RNA sequence. The
CC DNA mol. enables AGO to be expressed safely in plants, i.e.
CC transgenic plants develop normally, which imparts resistance to
CC fungal and bacterial pathogens (e.g. Verticillium dahliae,
CC Phytophthora infestans, Botrytis cinerea, Septoria nodorum,
CC Pseudocercospora herpotrichoides, Gaumannomyces graminis and
CC Erwinia carotovora) in the plants, specifically potatoes.
CC Potato tubers transformed with the DNA mol. were incubated with
CC 50000 c.f.v. E. carotovora for 3 days at 23 degrees C at high
CC humidity. The final number of bacteria/lesion for tubers that
CC expressed AGO at high level was 0.855-4.55 billion, compared to
CC 16.76 billion for controls.
SQ Sequence 1848 BP; 387 A; 592 C; 470 G; 399 T;

```

alignment\_scores:

```

Quality: 91.50      Length: 280
Ratio: 0.863       Gaps: 19
Percent Similarity: 37.857  Percent Identity: 24.286

```

alignment\_block:  
US-09-209-961-5 x T29485

Align seg 1/1 to: T29485 from: 1 to: 1848

```

4 PROArgAlaGArgTyrArgArgArgHisArgProArgSerHisLeucl 20
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
939 CTGGCCCGGGCTCGCTGCTCTCTCCCAATCTCGA..... 977

20 yGlnIleLeuArgArg.....ProTyrLeuV 30
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
978 ....ATATTCCGGTATCGAATGAAGTCATCTCGAGCCCTTGGTATC 1023

30 aHis.....ProArgHis 34
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1024 GACACCGCTGTTGACCTGCCCTCGGCTTGAACCTGCAGGACGACACAC 1073

35 ArgTyrArg.....Tyr.....ArgArgLysAsnG 43
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1074 CGCTACCTCGCTCCCGCATCACCTCTGCTGTCAGGACGAGGACAGG 1123

43 yIlePheAsnThrArgLeuSerArg..ThrPheGly...TyrThrVallys 58
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1124 CGCTTGGTTCGCCACCTCAACGAGACCTTTGGTGACTATTCCGGAAG 1173

59 ArgThrThrValThrThrProSerThrAlaValAspMetMetArgPheL 75
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1174 GCACAGAGCTGCTCAACACCAAGCTGGAGCAGTGG..... 1209

75 ysIleAspAspPheValProProGlyGly.....GlyThrAsnLysIle 89
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1210 ..GCCGAGAGGCGCTGCCCGCTGGCGGATTCCCAACACACCGCCTTG 1257

90 SerIleProPheGluTyrTyrArg..... 97
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1258 CTCATCCAGTAGAGAACTACCGCGACTGGATTGTCAACCAACAGTCGC 1307

98 .....IleArgLysValLysValGluPheT 106
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1308 GTACTCGGAACCTTCTCTCGACACTGCCGAGTAGCAGCTTCGATGT 1357

106 rpProCysSerProIleThrGlnGlyAspArgGlyValGlySerThrAla 122
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1358 GGGACCTTCTGCCCTTCACCCGAGGA.....TACGTT 1389

123 ValIleLeuAspAspAsnPheValThrLysAlaThrAlaLeuThrTyrAs 139
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1390 CACATCTCTCGACAAG.....GA 1406

139 pProTyrValAsnTyrSerSerArgHisThrIleProGlnProPheSert 156
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1407 CCGCTACCTTCAACCA.....TTGGCCT 1429

156 yHisSerArgTyrPheThrProLysProValLeuAspSerThrIleasp 172
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1430 ACGACCCCTCAGTACTCTC..... 1449

173 TyrPheGlnProAsnAsnLysArgAsnGlnLeuTrpLeu..... 185
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1450 .....AACGAGCTGGACCTGCTCGTCAAGC 1475

186 ....ArgLeuGlnThrSerGlyAsnValAspHisValGly...LeuGlyT 200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1476 TGGCGTACTCAACTGCGCGCCCAACATCTCCAACCTCGGTGCCATGACAGA 1525

200 hrAlaPhe.....GluAsnSerLysTyrAspGln 209
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1526 CCTACTTCGCTGGGAGACTATCCCGGTGATACCTCGCGTATGATGCC 1575

210 AspTyrAsnIleArgValThrMetTyrValGlnPhe 221
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1576 GATTGTAGC...GCTGGACTGAGTACATCCCGTAC 1608

```

```

alignment_scores:      Quality: 91.50      Length: 280
                       Ratio: 0.863      Gaps: 19
Percent Similarity: 37.857      Percent Identity: 24.286

alignment_block:
US-09-209-961-5 x Q03143  ..

Align seg 1/1 to: Q03143 from: 1 to: 2024

      4  ProArgArgArgTyrArgArgArgHisArgProArgSerHisLeuG1 20
          ||| ||| ||| ||| ||| ||| |||
     954  CTTGGCCGGGGCTCCGGTGTCTCTCCACAATCTCGA..... 992

      20 yGlnIleLeuArgArgArg.....ProTrpLeuV 30
          ||| ||| ||| ||| ||| ||| |||
     993  ....ATATTCGGTATCGGAATGAAGTCCATCTGGAGCCCTTGGTATC 1038

      30 alHis.....ProArgHis 34
          ||| ||| ||| ||| ||| |||
    1039  GACACCGTCGTTGACCTCCCGTCGCTTGAACCTGCAGGACCAACCAC 1088

      35 ArgTyrArg.....Trp...ArgArgLysAsnG1 43
          ||| ||| ||| ||| ||| |||
    1089  CGCTACCGTCCGCTCCCGCATCACCTCTGCTGGTGCAGGACAGGACAGG 1138

      43 yIlePheAsnThrArgLeuSerArg.ThrPheGly...TyrThrVallys 58
          | ::: :||| :||| ||| ||| ||| ||| ||| :|||

```

```

seq_name: N_Geneseq_36:X33246
seq_documentation_block:
ID   X33246 standard; DNA; 1015 BP.
AC   X33246;
CD   30-JUN-1999 (first entry)
DT   30-JUN-1999 (first entry)
DE   PEBP2 alpha A gene expression regulating DNA sequence #2.
KW   PEBP2 alpha A gene; expression; regulation; bone disease;
KW   osteoporosis; ds.
OS   Mus sp.
PN   WO9911787-A1.
PD   11-MAR-1999.
PF   02-SEP-1998; J03920.
PR   08-APR-1998; JP-114135.
PR   02-SEP-1997; JP-254250.
PR   15-OCT-1997; JP-299407.
PA   (SUMO ) SUMITOMO PHARM CO LTD.
PI   Fujiwara M, Harada H, Katsumata T, Nakatsuka M,
PI   Ogawa S, Tagashira S;
PI   WPI: 99-243621/20.
PT   DNA regulating expression of PEBP2 alphaA gene to produce r
PT   protein, useful as promoter for prevention of/and treatment
PT   diseases e.g. osteoporosis

```

PS Claim 1; Page 74-75; 118pp; Japanese.  
CC The present sequence represents DNA which participates in the regulation  
CC of expression of PEBP2 alpha A gene. The DNA produces a regulator  
CC protein with the activity of promoting bone formation and can serve as a  
CC promoter for prevention and treatment of bone diseases including  
CC osteoporosis.  
SQ Sequence 1015 BP; 163 A; 376 C; 327 G; 149 T;

alignment\_scores:  
Quality: 87.50 Length: 196  
Ratio: 1.029 Gaps: 11  
Percent Similarity: 43.367 Percent Identity: 25.510  
alignment\_block:  
US-09-209-961-5 x X33246 ..  
Align seg 1/1 to: X33246 from: 1 to: 1015

4 ProArgArgTyrArgArgHis.....ArgProAr 16  
|||||  
460 CCTCGAGGAGGCGGAGGAGGCGCGGCAAGCGAGCGCGCG 509  
16 gSerHisLeuGlyGlnIleLeuArgArgProTirLeuValHisProA 33  
|||  
510 AGCT.....GGGCAG.....CCGC 523  
33 rgHisArgTyrArgTirArg.....ArgLysasn 42  
|||  
524 GCTCGTGTGGCGTGGCGGACAGCGGAGGCGCACACGCGCGAGAT 573  
43 GlyIlePheAsnThrArgLeuSerArgThrPhe.....GlyTyrTh 56  
|||||  
574 GGACTGTGAACCTCGCGGGTCTCACTACCTACTGGAACCGGAGCGG 623  
56 rValLysArgThrValThrPro..... 65  
624 GCGCCAGCGCACCGAGGACAGCGGTGCCCGCCAGGCGCCCTACTGCAAGCT 673  
66 .....SerTirPalaValAspMetMetArgPhe 74  
674 GTTAACCTCAAGTCCCTCGCGGCTGGGCCAACACGCGCGCTCC 723  
75 LysIle.....AspAspPheValProProGlyGlyTh 86  
|||  
724 CGCTCGTAGACGCCACCGCGGACTCCCAAGCCTCCGCTGGAGATCC 773  
86 rAsnLysIleSerIleProPheGluTyrTyrArgIleArgLysValLysV 103  
|||  
774 GCGCTCTCGGCTGCCCGGCTACCGGACTCTGTCGGTCTCCAG 823  
103 alGluPheTirProCysSerProIleThrGlnGlyAspArgGlyValGly 119  
|||  
824 TC.....GCGCGCCCGCCAGCAGACTCCGCGCGGACACGCGCT 861  
120 SerThrAlaValIleLeuAspAspAsnThrValThrLysAlaThrAlaLe 136  
|||  
862 TCCTCGAACCTTGAAT.....TCTACCTCTCTGTCCTCCCTCCATCCAT 905  
136 uThrTyrAspProTyrValAsnTyrSerSerArgHisThrIleProGlnP 153  
|||  
906 CCTCTTTCGCGCC.....CGCTCTCGGCTCCAC.....CCCTCG 940  
153 ropHeSerTyrHisSerArgTyrPheThrProLysPro 165  
|||  
941 ATTTCCTCTCTCGCGCCCTTCCTCCCTCCCTCC 978

seq\_name: N\_Geneseq\_36:062817

seq\_documentation\_block:  
ID Q62817 standard; DNA; 1738 BP.  
AC Q62817:  
DT 10-NOV-1994 (first entry)

DE Genomic sequence encoding human alpha 1 B adrenergic receptor.  
KW Alpha 1B; adrenergic receptor; antagonist; neuroreceptor; ss.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT cds 124..1683  
FT /\*tag= a  
PN WO9410989-A.  
PD 26-MAY-1994.  
PF 12-NOV-1993; U10950.  
PR 13-NOV-1992; US-973867.  
PA (SYN-) SYNAPTIC PHARM CORP.  
PI Branchek TA, Chiu G, Forray CC, Gluchowski C, Hartig PR,  
PI Wetzel JM.  
DR WPI: 94-183130/22.  
DR P-PSDB; K52831.  
PT Treating benign prostatic hyperplasia - with cpds. which bind  
PT selectively to alpha1C adrenergic receptor to inhibit contraction  
PT of prostatic tissue  
PS Example: Page 46-51; 69pp; English.  
CC The inventors claim a method of treating benign prostatic  
CC hyperplasia which comprises administering a therapeutically  
CC effective amount of a compound which binds to a human alpha-1C  
CC adrenergic receptor with a binding affinity greater than ten-fold  
CC higher than the binding affinity with which the compound binds to a  
CC human alpha-1A adrenergic receptor or a human alpha-1B adrenergic  
CC receptor. The example concerns a protocol for the determination of  
CC the potency of alpha-1 antagonists. Cell lines were prepd. by  
CC transfecting the cloned cDNA or cloned genomic DNA or constructs  
CC contg. both genomic DNA and cDNA encoding human alpha-adrenergic  
CC receptors. For alpha-1A, the entire coding region and 5' and 3'  
CC UTRs were cloned into the BamHI and ClaI sites of the  
CC polylinker-modified eukaryotic expression vector pCEXV-3, called  
CC EXJ.HR. For alpha-1B, the entire coding region and 5' and 3' UTRs  
CC were cloned into the EcoRI site of pCEXV-3 eukaryotic expression  
CC vector. For alpha-1C, the entire coding region and 5' and 3'  
CC UTRs were cloned into the KpnI site of the polylinker-modified  
CC pCEXV-3-derived eukaryotic expression vector, EXJ.RH.  
SQ Sequence 1738 BP; 308 A; 610 C; 495 G; 325 T;

alignment\_scores:  
Quality: 90.00 Length: 127  
Ratio: 1.385 Gaps: 8  
Percent Similarity: 51.181 Percent Identity: 28.346  
alignment\_block:  
US-09-209-961-5 x Q62817/rev ..  
Align seg 1/1 to reverse of: Q62817 from: 1 to: 1738  
4 ProArgArgTyrArgArgArgHisArgProArgSerHisLeuG1 20  
|||||  
1268 CCCAGCGACGCGCGCGGCTCGCGCGCGCGCGCGCG..... 1227  
20 yGlnIleLeuArgArgProTirLeuValHisProArgHisArgTyrA 37  
|||  
1226 .....CGGCACTGG.....CACCGGAGGATCGGCACGA 1199  
37 rGTrpArgArgLysAsnGlyIlePheAsn.....ThrArgLeuSer 50  
|||  
1198 AAGCGCGCTTGAAGTCTCTGCGGACATGGGTAGATGGGTGAGG 1149  
51 ArgThrPheGlyTyrThrValLysArgThrValThrProSerTr 67  
|||  
1148 CAGCTGTGAAGTAGCCAGCAGAACACACCTTGAACACGCGCGTGG 1099  
67 palavalAspMetMetArgPheLysIleAspAspPheValProGly. 83  
|||  
1098 GGGC.....TTCAGGTGGAGAACAAAGGAGCGCAAGCGGTA 1064  
84 .....GlyGlyThrAsnLysIleSerIlePro..... 92  
1063 GAGCGATGAAGAGGTAGCCAGCACACAGATGACATACCGACCACATG 1014



93 .....PheLUtYrYrArGileArGLysVallys..... 102  
 1013 CCAACGCTCTAGCTCTTTCTTTCTGGAGAACTTAAAGATTGAC 964  
 103 .....ValGLuPhe.TrpProCysSerPro 110  
 963 AGCTATGAAGAACTCTGGGGTTGTGGCCT 935

seq\_name: N\_Geneseq\_36:X20631

seq\_documentation\_block:

ID X20631 standard; DNA; 722 BP.  
 AC X20631;  
 DE 05-MAY-1999 (first entry)  
 DE Polynucleotide sequence from the genome of Treponema pallidum.  
 KW Treponema pallidum infection; syphilis; Borrelia infection; animal;  
 KW enzyme production; ds.  
 OS Treponema pallidum.  
 PN WO9859034-A2.  
 PD 30-DEC-1998.  
 PF 23-JUN-1998; U13041.  
 PR 24-JUN-1997; US-050667.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Fraser CM;  
 DR WPI; 99-081273/07.  
 PT New isolated Treponema pallidum nucleic acids - used to develop  
 PT products for the detection, diagnosis, characterisation, prevention  
 PT and therapy of T. pallidum infections, particularly syphilis  
 PS Claim 1; Page 763; 1150pp; English.  
 CC X20500-21243 represent polynucleotide sequences from the genome of  
 CC Treponema pallidum. The sequences can be used for detection,  
 CC diagnosis, characterisation, prevention and therapy for T. pallidum  
 CC infections, particularly syphilis. They can also be used for detecting  
 CC diseases related to Borrelia infections in animals, and for the  
 CC production of biosynthetic products such as enzymes.  
 SQ Sequence 722 BP; 183 A; 152 C; 221 G; 166 T;

alignment\_scores:

Quality: 83.50 Length: 187  
 Ratio: 0.949 Gaps: 12  
 Percent Similarity: 47.059 Percent Identity: 26.738

alignment\_block:

US-09-209-961-5 x X20631 ..  
 Align seg 1/1 to: X20631 from: 1 to: 722  
 1 MetThrTyProArgArgTyArgArgArgHisArgProArg.. 16  
 27 FTGAGGTATGCGGCGAAAGTTCTCTCGAAGCAGG...AGACCGAAGAC 73  
 17 .....SerHisLeuGlyGlnIleLeuArgA 25  
 74 ACCAGGGGAAGAAAAGTGGGAGTACGAGACGACCCAGCGTTACTAA 123  
 25 tGArgProTrpLeuValHisProArgHisArgTyArgTrpArgLys 41  
 124 GA.....TGGTGGCGCTCTCGCGTCATTCAGGATTTGGGAGGACG 167  
 42 AsnGlyIlePheAsnThrArgLeuSerArgThrPheGlyTyThrVally 58  
 168 GGG.....AGATTAAGT.....TTGAA 184  
 58 sArgThrThrVal.....ThrThrProSerTrpA 68  
 185 GCAGTCGAGGGTGACAGTACGTTGGCGGATCGCGGAGTTCTTCATGTT 233  
 68 laValAspMetMetArgPheLysIleAspAsp..... 78  
 234 T....GACAGCGAGGANTACAGATTACGACNACGTAAGGTTTCAGGTATG 280

79 ...PheValProProGlyGlyGlyThrAsnLysIleSerIleProPheG1 94  
 281 AAGTTTGTCCCA.....GTTGGGTTCCTCATGA 309  
 94 uTyTyArgIleArgLysValGluPheTrpProCysSerProI 111  
 310 ATTAAGAGGATTGCAAGGAGAGTTTCACTCTGTGGAAGACTCCCGC 359  
 111 leThrGlnGlyAspArgGlyValGlySerThrAlaValIleLeuAspAsp 127  
 360 TTACGGAGATACCAACGCCCTT.....AAGACAATGCTCACTGGAGAT 403  
 128 AsnPheValThrLysAlaThrAlaLeuThrTyAspProTyValasNTY 144  
 404 AGTTTCTGCACGTAAGGTAAGCAGCAGCAGCCGCGACGA....CC 449  
 144 rSerSerArgHis.....ThrIleProGlnProPheSerTyHis 158  
 450 TTGTGTAGACACGCGTGGGTACCGGTTACCCAGCCGTTTGTGTCGGAC 499  
 158 erArgTyPhe 161  
 500 GCAGAGGCTTC 510  
 seq\_name: N\_Geneseq\_36:X18977  
 seq\_documentation\_block:  
 ID X18977 standard; DNA; 3722 BP.  
 AC X18977;  
 DE 13-MAY-1999 (first entry)  
 DE Non-B, non-C, non-G hepatitis virus gene DNA sequence SEQ ID NO:62.  
 KW Hepatitis virus; non-B non-C non-G hepatitis virus; PCR primer;  
 KW virion; detection; screening; infection; blood transfusion; ss.  
 OS Hepatitis virus.  
 PN WO9905282-A1.  
 PD 04-FEB-1999.  
 PF 27-JUL-1998; J03340.  
 PR 13-MAR-1998; JP-082962.  
 PR 25-JUL-1997; JP-233246.  
 PR 09-OCT-1997; JP-314196.  
 PA (TAMU/) TAMURA R.  
 PI Nishizawa T, Okamoto H;  
 DR WPI; 99-142937/12  
 PT Gene isolated from non-B non-C non-G hepatitis DNA virus - and its  
 PT expression products, useful for diagnosis and treatment of hepatitis  
 PT and screening of blood for transfusion.  
 PS Claim 37; Page 102-103; 113pp; Japanese.  
 CC The present sequence represents a gene which is contained in a non-B  
 CC non-C non-G hepatitis DNA virus. It is 3500-4000 bases in length and  
 CC contains two overlapping open reading frames (ORF). It is obtained by  
 CC polymerase chain reaction (PCR) amplification using primer. The gene can  
 CC be used for the production of vaccines for prevention and treatment of  
 CC non-B non-C non-G hepatitis infection. Diagnosis of such infection, and  
 CC screening of blood (e.g. intended for transfusion) for the presence of  
 CC the virus, by using the virion or antigenic peptides as reagents for  
 CC detection of antibodies to the virus, or by direct detection of the gene  
 CC using PCR with primers derived from the gene sequence.  
 SQ Sequence 3722 BP; 1092 A; 1029 C; 825 G; 776 T;

alignment\_scores:

Quality: 91.50 Length: 232  
 Ratio: 0.934 Gaps: 14  
 Percent Similarity: 42.241 Percent Identity: 25.862

alignment\_block:

US-09-209-961-5 x X18977 ..  
 Align seg 1/1 to: X18977 from: 1 to: 3722  
 4 ProArgArgArgTy.....ArgArgArgHisArgProAR 16  
 623 CCGAGCGCGCGTGGAGAGGTGGCGCGCGTGAAGAGACTACCGAG 672







OM of: US-09-209-961-5 to: Issued Patents\_NA:\* out\_format : pfs  
Date: Dec 31, 1999 12:26 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-1998 CompuGen Ltd.

Command line parameters:  
-MODEL=framed.p2n.model -DEV=xlp  
-Q/cgn2\_1/USPTO.spool/US09209961/runat\_22121999\_101043\_17921/app\_query.fasta.1  
-DB=issued\_patents\_NA -OFF=fastap -SUFFIX=rni -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-GAPOP=6.000 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blowsum62  
-TRANS=human40.cdi -LIST=45 -DOCAIGN=200 -THR\_SCORE=escore  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09209961  
-NCPU=6 -ICPU=3 -NO\_XLPXY -WAIT -THREADS=1

Search information block:  
Query: US-09-209-961-5  
Query length: 233  
Database: Issued Patents\_NA\*  
Database sequences: 195662  
Database length: 53930012  
Search time (sec): 88.010000

score_list:	Strd Orig	ZScore	Escore	Len	Documentation
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-333-802-1 +	91.50	154.91	0.3121	1848	1
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-307-499-53 +	85.00	148.37	0.7227	1023	1
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-334-998-3 -	87.00	146.26	0.9466	1738	1
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-228-932-3 -	87.00	146.26	0.9466	1738	1
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-468-939-3 -	87.00	146.26	0.9466	1738	1
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-406-855A-3 -	87.00	146.26	0.9466	1738	1
/cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US95-04203-3	87.00	146.26	0.9466	1738	1
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-07-814-220-23 -	67.00	142.26	1.58	72	1
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-07-812-421-23 -	67.00	142.26	1.58	72	1
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-940-245-1 +	83.50	141.72	1.69	1377	1
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-226-486-1 +	83.50	141.72	1.69	1377	1
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-403-852D-14 +	75.50	137.67	2.85	474	1
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-910-973-21 +	80.50	135.86	3.60	1333	1
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-910-973-12 +	80.00	135.41	3.81	1268	1
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-314-309A-5 +	77.50	130.76	6.91	1209	1
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-07-829-461A-8 -	65.00	129.41	8.22	150	1
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-072-282-1 -	72.50	129.20	8.45	572	1
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-470-892-1 -	71.50	128.68	9.03	501	1
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-314-309A-2 +	77.50	128.11	9.72	1513	1
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-314-309A-3 +	77.50	127.85	10.04	1546	1
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-07-829-461A-8 -	64.00	127.32	10.74	150	1
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-831-570-2 -	72.50	124.61	15.22	843	1
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-722-001-24 -	76.00	124.57	15.30	1567	1
/cgn2_6/ptodata/1/ina/backfiles.seq:5248599-1 +	77.50	122.58	19.75	2414	1
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-974-546-4 +	69.50	122.52	19.90	593	1
/cgn2_6/ptodata/1/ina/backfiles.seq:5244792-2 +	79.00	121.40	22.96	3472	1
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-923-692C-5 +	74.00	121.32	23.21	1450	1
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-184-237-5 +	74.00	121.32	23.21	1450	1
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-482-920-5 +	74.00	121.32	23.21	1450	1
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-343-428-1 +	76.00	121.31	23.25	2084	1
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-535-230A-1 +	72.50	120.48	25.85	1195	1
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-042-747A-7 +	77.00	119.19	30.49	2943	1
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-076A-10 +	76.00	117.89	36.01	2754	1
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-299-498A-60 -	58.00	117.76	36.63	117	1
/cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US95-10813-60 -	58.00	117.76	36.63	117	1
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-265-429A-4 -	71.50	117.68	36.99	1269	1
/cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US95-09069-4 -	71.50	117.68	36.99	1269	1
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-307-499-1 +	85.00	117.26	39.06	14176	1
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-307-499-14 +	85.00	117.26	39.06	14176	1
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-157-363A-9 +	77.50	117.13	39.72	3826	1
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-377-825-1 +	77.50	117.13	39.72	3826	1
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-922-175-1 +	77.50	117.13	39.72	3826	1

/cgn2\_6/ptodata/1/ina/PCTUS9\_COMB.seq:PCT-US93-11405A-9 + 77.50 117.13 39.72  
/cgn2\_6/ptodata/1/ina/PCTUS9\_COMB.seq:PCT-US93-11405-9 + 77.50 117.13 39.72  
/cgn2\_6/ptodata/1/ina/5D\_COMB.seq:US-08-765-875-1 + 71.50 116.58 42.61 1393  
seq\_name: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:US-08-333-802-1  
seq\_documentation\_block:  
: Sequence 1, Application US/08333802  
: Patent No. 5516671  
: GENERAL INFORMATION:  
: APPLICANT: Lawrence, Ellen B.  
: APPLICANT: Levine, Elaine B.  
: APPLICANT: Shah, Dilipkumar M.  
: TITLE OF INVENTION: Method of Controlling Plant Pathogens  
: NUMBER OF SEQUENCES: 4  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Grace L. Bonner, Monsanto Company, BB4F  
: STREET: 700 Chesterfield Parkway NO. 5516671th  
: CITY: St. Louis  
: STATE: Missouri  
: COUNTRY: USA  
: ZIP: 63198  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/333,802  
: FILING DATE:  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 08/161041  
: FILING DATE: 24-NOV-1993  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Bonner, Grace L.  
: REGISTRATION NUMBER: 32,963  
: REFERENCE/DOCKET NUMBER: 38-21(10663)A  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (314)537-7286  
: TELEFAX: (314)537-6047  
: INFORMATION FOR SEQ ID NO: 1:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 1848 base pairs  
: TYPE: nucleic acid  
: STRANDEDNESS: double  
: TOPOLOGY: linear  
: MOLECULE TYPE: DNA (genomic)  
: FEATURE:  
: NAME/KEY: CDS  
: LOCATION: 16..1833  
: US-08-333-802-1  
alignment\_scores:  
Quality: 91.50 Length: 280  
Ratio: 0.863 Gaps: 19  
Percent Similarity: 37.857 Percent Identity: 24.286  
alignment\_block:  
US-09-209-961-5 x US-08-333-802-1 ..  
Align seq 1/1 to: US-08-333-802-1 from: 1 to: 1848  
4 ProArgArgArgTyrArgAgAgAgAgHisArgProArgSerHisLeuG1 20  
|||||  
939 CCGGGCGGGGCTCGCTGCTCTCCCAATCTCGA..... 977  
20 yGinileLeuArgArg.....Protrileuv 30  
|||||  
978 .....ATATTCGGGTATCGGATGATCCATCTGGAGCCCTTGGTATC 1023  
30 alHis.....ProArgHis 34

```
1024 GACACCGTCGTTGACCTGCCCGTGGCTTGAACCTGCAGGACCAACAC 1073
1035 ArgTyrArg.....Trp...ArgArgLysAsnG1 43
1074 CGCTACCGTCGCTCCCGCATCACCTCTGCTGGTCAGGACAGGACAGG 1123
43 yilePheAsnThrArgLeuSerArg, ThrPheGly...TyrThrValLys 58
1124 CCGCTTGGTGGCCACCTTCAACGAGACTTGGTGACTATTCCGAAAG 1173
59 ArgThrThrValThrThrProSerThrAlaValAspMetMetArgPheL 75
1174 GCACAGAGCTGCTCAACACCAAGCTGGAGCAGTGG..... 1209
75 ysileAspAspPheValProProGlyGly.....GlyThrAsnLysile 89
1210 ..GCCGAGAGGCCGCTGCCCGTGGCGGATTCCACAACACCAACCGCTTG 1257
90 SerileProPheGluTyrTyrArg..... 97
1258 CTCATCCAGTACGAGACTACCGCGCTGGATGTCAACCAACACGTCGC 1307
98 .....ileArgLysValLysValGluPheL 106
1308 GTACTCGGAACCTTCCTCGACACTGCCGGAGTAGCCAGCTTCGATGT 1357
106 rpProCysSerProIleThrGlnGlyAspArgGlyValGlySerThrAla 122
1358 GGGACCTTCTGCCCTTCACCGAGCA.....TACGTT 1389
123 valIleLeuAspAspAsnPheValThrLysAlaThrAlaLeuThrTyrAs 139
1390 CACATCTCTCGAAG.....GA 1406
139 pProTyrValAsnTyrSerArgHisThrIleProGlnProPheSerT 156
1407 CCCCCTACCTTCACCA.....TTCGCT 1429
156 yHisSerArgTyrPheThrProLysProValLeuAspSerThrIleAsp 172
1430 ACGACCCCTCAGTACTCTC..... 1449
173 TyrPheGlnProAsnAsnLysArgAsnGlnLeuTrpLeu..... 185
1450 .....AACGAGCTGGACCTGCTCGGTACAGC 1475
186 ....ArgLeuGlnThrSerGlyAsnValAspHisValGly...LeuGlyT 200
1476 TGGCGCTACTCAACTGGCCGCAACATCTCCAACCTCGGTGATGATGCC 1525
200 hrAlaPhe.....GluAsnSerLysTyrAspGln 209
1526 CCTACTTCGTTGGGAGACTATCCCGGTGATACCTCGCGTATGATGCC 1575
210 AspTyrAsnIleArgValThrMetTyrValGlnPhe 221
1576 GATTGAGC...GCCTGGAGCTGAGTACATCCCGTAC 1608
```

seq\_name: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.us-08-307-499-53

```
seq_documentation_block:
; Sequence 53, Application US/08307499
; Patent No. 5651972
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Vi uela, Eladio
; APPLICANT: Gibbs, E.P.J.
; TITLE OF INVENTION: Use of Recombinant Swine Poxvirus as a
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRES:
; ADDRESSEE: David R. Saliwanchik
```

```
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: Florida
COUNTRY: U.S.A.
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/307,499
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/908,241
FILING DATE: 1-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/908,630
FILING DATE: 29-JUN-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/342,212
FILING DATE: 21-APR-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: UF35.1.FWCC1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 53:
SEQUENCE CHARACTERISTICS:
LENGTH: 1023 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-08-307-499-53
```

alignment\_scores: Quality: 85.00 Length: 126  
Ratio: 1.545 Gaps: 6  
Percent Similarity: 43.651 Percent Identity: 27.778

```
alignment_block:
US-09-209-961-5 x US-08-307-499-53 ..
Align seg 1/1 to: US-08-307-499-53 from: 1 to: 1023
38 TTPArgArg.....LysAsnGlyIlePheAsnThrArgLeuSerAr 51
418 TGGACACGATTCCACCCCAAGATAAACCTTAGTAAGTTAGATCAGC 467
51 gThrPheGlyTyrThrValLysArgThrValThrProSerTrpA 68
468 TACATTTCCA...AAGGTAGGAATGTAACTGTAAATCACAACACTGG 514
68 laValAspMetMetArgPheLysIleAspAspPheValProGlyGly 84
515 AACGTGTTATGAACATCTT.....GGA 537
85 GlyThrAsnLysIleSerIleProPhe.....GluTyrTyrArgL 98
538 TCATGTGTAACATTACATGCTCCGTTTACAGCGGATGATATATAAAT 587
98 eArgLysValLysValGlu.....PheTrpProCysSerProIleThrG 113
588 TTCTAAGGGATATATAGATAAGCCAGTTAAGCTACTGTACAGTTACAG 637
113 lngLysAspArgGlyValGlySerThrAlaValIleLeuAspAsnPhe 129
```

```
638 GAATTGAAGAGGAGATACATACATTGATATGCACATTTGTATAATCAT 687
130 ValThrLysAlaThrAlaLeu..... 136
688 TATCGTCGTCGCTGCTTTAAATGGTTAATCATCGAGGACTTTCCTCC 737
137 .....ThrTyrAspProTyrValAsn 143
738 GGACATATCGTTATGATCCGTCAGTAAAT 765
```

seq\_name: /cgn2\_6/ptodata/1/lna/5A\_COMB.seq:US-08-334-698-3

```
seq_documentation_block:
; Sequence 3, Application US/08334698
; Patent No. 5556753
; GENERAL INFORMATION:
; APPLICANT: Jonathan A. Bard et al.
; TITLE OF INVENTION: DNA Encoding Human Alpha 1 Adrenergic
; TITLE OF INVENTION: Receptors and Uses Thereof
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: COOPER & DUNHAM
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; FILING DATE:
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/952,798
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 376901
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 664-0525
; TELEX: (212) 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1738 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: N
; ANTI-SENSE: N
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 124..1683
; OTHER INFORMATION:
; US-08-334-698-3
```

alignment\_scores: Quality: 87.00 Length: 128  
Ratio: 1.318 Gaps: 9  
Percent Similarity: 51.562 Percent Identity: 28.125

alignment\_block:

US-09-209-961-5 x US-08-334-698-3/rev ..

Align seg 1/1 to reverse of: US-08-334-698-3 from: 1 to: 1738

```
4 ProArgArgTyrArgArgArgHisArgProArgSerHisLeuG1 20
1268 CCCAGCGACGCGCGCGCGCTCGCGCGCGCGCGCGCG..... 1227
20 yGlnIleLeuArgArgArgProTyrLeuValHisProArgHisArgTyrA 37
1226 .....CGGCACTGG.....CACCCGAGGATGGCGCAGA 1199
37 rGTpArgArgLysAsnGlyIlePheAsn.....ThrArgLeu 49
1198 AA...GGCGCTTGAACCTCTGCTGGAGCATGGGTAGATGATGGGGTGT 1152
50 SerArgThrPheGlyTyrThrValLysArgThrThrValThrThrProse 66
1151 AGCAGCTGTGAAGTAGCCAGCCAGACACACCTTGAACACGGCGTC 1102
66 rTTPAlaValAspMetMetArgPheLysIleAspPheValProProG 83
1101 GGGGGGC.....TTCAAGGTGGAGAACAAAGGAGCAAGCG 1067
83 ly.....GlyGlyThrAsnLysIleSerIlePro..... 92
1066 GTAGAGCGATGAAGAGGGTAGCCAGCACAGATGAACATACCCACCACA 1017
93 .....PheGluTyrTyrArgIleArgLysValLys..... 102
1016 ATGCCCAACGCTTAGCTGCTTTCTTCTCCTGGAGAACTTAAAAAGTTT 967
103 .....ValGluPhe.TrpProCysSerPro 110
966 GACAGCTATGAAGTCTCTGGGGTGTGGCCCT 935
```

seq\_name: /cgn2\_6/ptodata/1/lna/5A\_COMB.seq:US-08-228-932-3

```
seq_documentation_block:
; Sequence 3, Application US/08228932
; Patent No. 5578611
; GENERAL INFORMATION:
; APPLICANT: Charles Gluchowski, Carlos C. Forray, George Chiu,
; APPLICANT: Theresa A. Brancheck, John M. Wetzel and Paul R. Hartig
; TITLE OF INVENTION: USE OF ALPHA-1C SPECIFIC COMPOUNDS TO TREAT BENIGN
; TITLE OF INVENTION: PROSTATIC HYPERPLASIA
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: COOPER & DUNHAM
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/228,932
; FILING DATE: 13-APR-1994
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 41878-B/JPW/TEP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 664-0525
; TELEX: (212) 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1738 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
```

seq\_name: /cgn2\_6/ptodata/1/1na/5C\_COMB.seq:US-08-406-855A-3



```

seq_documentation_block:
; Sequence 3, Application US/08406855A
; Patent No. 5861309
; GENERAL INFORMATION:
; APPLICANT: Jonathan A. Bard et al.
; TITLE OF INVENTION: DNA Encoding Human Alpha 1 Adrenergic
; TITLE OF INVENTION: Receptors and Uses Thereof
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/406,855A
; FILING DATE: 21-AUG-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 41337-A-PCT-US/JPW/KDB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0526
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1738 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: N
; ANTI-SENSE: N
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 124..1683
; OTHER INFORMATION:
;
US-08-406-855A-3

```

```

alignment_scores:
    Quality: 87.00      Length: 128
    Ratio: 1.318        Gaps: 9
    Percent Similarity: 51.562    Percent Identity: 28.125

alignment_block:
    US-09-209-961-5 x US-08-406-855A-3/rev ..

Align seg 1/1 to reverse of: US-08-406-855A-3 from: 1 to: 1738

      4  ProArgArgGTYrArgArgArgHisArgProArgSerHisLeuGl 20
          |||||
      1268  CCAGCGCAGCGCGCGCGCGCTCGCGCGCGCGCGCG..... 1227
          |||||

      20  yGlnlleLeuArgArgArgProTrpLeuValHisProArgHisArgTYrA 37
          |||||
      1226  .....CGGCACATGG.....CACCCGAGGATGCCACGA 1199
          |||||

      37  rgTrpArgArgLysAsnGlnYillePheAsn.....ThrArgLeu 49
          :|||
          :|||
      1198  AA...CGCGCTTGAACTCCTTGCCTGGAGCATGGTAGATGGGGTTG 1152
          :|||

      50  SerArgThrPheGlyTYrThrValLysArgThrThrValThrProSe 66
          :|||
          :|||
      1151  AGCGACTGCTTGAAGTAGCCGAGCCAGCAACACACCTTGAACAGCGCTC 1102
          :|||
          :|||

```

```

66 rTrpAlaValAspMetMetargPhelysileAspAspPhValProProg 83
| : : :
1101 GGGGGC.....TTCAGGTGGAGAACAAAGGAGCCAAAGC 1067

83 ly.....GlyGlyThrAsnLysIleSerIlePro..... 92
|| : : : : : : : : : : : : : : : : : : : :
1066 GTAGACGCGATGAAGAAGGTAGCCAGCACAAAGATGAACATACCAGCACCA 1017

93 .....PheGluTyrtYrArgIleArgLysValLys..... 102

1016 ATGCCCAACGCTTACGTCTTCTCTTTTCCCTCGGAGAACTTAAAAAGTTT 967

103 .....ValGluPhe.TipProCysSerPro 110
: : : : : : : : : : : : : : : : : : : :
966 GACAGCTATGGAACCTCGGGGTGTGGCCCT 935

seq_name: /cgn2_6/ptodata/1/1na/PCrTUS9_COMB.seq: PCT-US95-04203-3

seq_documentation_block:
; Sequence 3, Application PC/TUS9504203
; GENERAL INFORMATION:
; APPLICANT: Charles Gluchowski, Carlos C. Forray, George
; APPLICANT: Chiu, Theresa A. Brancheh, John M. Wetzel and P
; TITLE OF INVENTION: USE OF ALPHA-1C SPECIFIC COMPOUNDS TO
; TITLE OF INVENTION: TREAT BENIGN PROSTATIC HYPERPLASIA
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: COOPER & DUNHAM LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04203
; FILING DATE: 4-APR-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 41878-D-PCT/JPW/AGL
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1738 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: N
; ANTI-SENSE: N
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 124..1683
; OTHER INFORMATION:
; PCT-US95-04203-3

```

```

alignment_scores:
  Quality: 87.00      Length: 128
  Ratio: 1.318      Gaps: 9
  Percent Similarity: 51.562      Percent Identity: 28.125

alignment_block:
  US-09-209-961-5  x  PCT-US95-04203-3/rev  ..

```

TELEPHONE: 703-391-2310

```

; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "synthetic DNA"
US-07-812-421-23

```

```

alignment_scores:
  Quality: 67.00      Length: 24
  Ratio: 3.350       Gaps: 0
  Percent Similarity: 83.333  Percent Identity: 54.167

alignment_block:
US-09-209-961-5 x US-07-812-421-23/rev ..
Align seg 1/1 to reverse of: US-07-812-421-23 from: 1 to: 72

```

```

4 ProArgArgTyrArgArgArgHisArgProArgSerHisLeuG1 20
|||||  |||  :|||  |||  |||  |||  |||  |||  |||  |||
72 CCGAGCGCGGCTGCGCAGCTACTGCGGCGACCGCAGCGCGTGGC 23

```

```

20 yGlnIleLeuArgArgPro 27
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
22 GCAGCTACTGCGGCGCGCGC 1

```

```
seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-940-245-1

```

```

seq_documentation_block:
; Sequence 1, Application US/07940245
; Patent No. 5350842
; GENERAL INFORMATION:

```

```

; APPLICANT: No. 5350842gard, Michael V
; TITLE OF INVENTION: Methods and Compositions for the
; TITLE OF INVENTION: Preparation of Treponema Pallidum Antigens, Including
; TITLE OF INVENTION: Purified Antigens and DNA Sequences Coding Therefor
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: Denise L. Mayfield
; STREET: 600 Congress Avenue, Suite 2300
; CITY: Austin
; STATE: TX

```

```

; COUNTRY: USA
; ZIP: 78701

```

```

; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/07/940,245
; FILING DATE: 31-AUG-1992

```

```

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: US 235351
; FILING DATE: 23-AUG-1988

```

```

; PRIOR APPLICATION DATA: US 913724
; APPLICATION NUMBER:

```

```

; FILING DATE: 30-SEP-1986
; ATTORNEY/AGENT INFORMATION:

```

```

; NAME: Denise L. Mayfield
; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: 512-320-7200
; TELEFAX: 512-474-7577

```

```

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 1377 base pairs
; TYPE: nucleic acid

```

```

; STRANDEDNESS: double
; TOPOLOGY: linear

```

```
US-07-940-245-1

```

```
alignment_scores:

```

```

Quality: 83.50      Length: 184
Ratio: 0.960       Gaps: 12
Percent Similarity: 47.283  Percent Identity: 25.543

```

```

alignment_block:
US-09-209-961-5 x US-07-940-245-1 ..

```

```
Align seg 1/1 to: US-07-940-245-1 from: 1 to: 1377

```

```

1 MetThrTyrProArgArgTyrArgArgArgHisArgProArg... 16
:::  |||  ::|||  |||  |||  |||  |||  |||  |||  |||
311 TTGAGTATTGGCGAAAGTTCTCTCGAAGCAGG...AGACCGAAGAC 357

```

```

17 .....SerHisLeuGlyGlnIleLeuArgA 25
:::  |||  |||  |||  |||  |||  |||  |||  |||  |||
358 ACAGGGGAAGAAAAGTGGGAGTACGAGACTGCCCAAGCGTTACTAA 407

```

```

25 rgArgProTrpLeuValHisProArgHisArgTyrArgTrpArgLys 41
||  |||  |||  |||  |||  |||  |||  |||  |||  |||
408 GA.....TGGTGGTGGCTCTCGCTCATTTTCAGGATTGGGAGAGGACG 451

```

```

42 AsnGlyIlePheAsnThrArgLeuSerArgThrPheGlyTyrThrVally 58
:::  |||  |||  |||  |||  |||  |||  |||  |||  |||
452 GGG.....AGATTAAGT.....TTGAA 468

```

```

58 sargThrThrVal.....ThrThrProSerTrpA 68
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
469 GCAGTCGAGGGTGCAGTACCGTTGGCGGATCGCGGAGTTCCTTCATGTT 517

```

```

68 laValAspMetMetArgPheLysIleAspasp..... 78
||  |||  |||  |||  |||  |||  |||  |||  |||  |||
518 T....GACAGCGGAGGAATACAAAGATTACGACGTAAAGGTTACCGGTATG 564

```

```

79 ....PheValProGlyGlyGlyThrAsnLysIleSerIleProPheG1 94
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
565 AAGTTTGTGCCA.....GTTGCGGTTCTTCATCA 593

```

```

94 uTyrTyrArgIleArgLysValLysValGluPheTrpProCysSerProI 111
|  |||  |||  |||  |||  |||  |||  |||  |||  |||
594 ATTAAAGGGATTGCAAGGAGAGATTTCACCTCGTGGGAAGACTCCCGCG 643

```

```

111 leThrGlnGlyAspArgGlyValGlySerThrAlaValIleLeuAspasp 127
::|  |||  |||  |||  |||  |||  |||  |||  |||  |||
644 TTACGGAGATACCAACGCGCTT.....AAGACAATGCTCCTAGAGAT 687

```

```

128 AsnPheValThrLysAlaThrAlaLeuThrTyrAspProTyr...Valas 143
::|  |||  |||  |||  |||  |||  |||  |||  |||  |||
688 AGTTTTTCTGCACGTAAGGTAAGCAGCATGGAGCGCCGACGACCTTGT 737

```

```

143 nTyrSerSerArgHisThrIleProGlnProPheSerTyrHisSerArgT 160
:::  |||  |||  |||  |||  |||  |||  |||  |||  |||
738 GGTAGACACGGTGGGTACCGTC.....TACCACAGCCGTT 772

```

```

160 yr 160
773 TT 774

```

```
seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-226-486-1

```

```
seq_documentation_block:

```

```

; Sequence 1, Application US/08226486
; Patent No. 5681934
; GENERAL INFORMATION:

```

```

; APPLICANT: No. 5681934gard, Michael V
; TITLE OF INVENTION: Methods and Compositions for the

```

```

; TITLE OF INVENTION: Preparation of Treponema Pallidum Antigens, Including
; TITLE OF INVENTION: Purified Antigens and DNA Sequences Coding Therefor

```

```

; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: Denise L. Mayfield
; STREET: 600 Congress Avenue, Suite 2300

```

```

; CITY: Austin
; STATE: TX

```

```

: COUNTRY: USA
: ZIP: 78701
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/226,486
: FILING DATE: 12-APR-1994
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/940,245
: FILING DATE: 31-AUG-1992
: APPLICATION NUMBER: US 235351
: FILING DATE: 23-AUG-1988
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 913724
: FILING DATE: 30-SEP-1986
: TELEPHONE: 512-320-7200
: TELEFAX: 512-474-7577
: NAME: Denise L. Mayfield
: ATTORNEY/AGENT INFORMATION:
: TELECOMMUNICATION INFORMATION:
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1377 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: US-08-226-486-1

alignment_scores:
  Quality: 83.50      Length: 184
  Ratio: 0.960        Gaps: 12
  Percent Similarity: 47.283  Percent Identity: 25.543

alignment_block:
  US-09-209-961-5 x US-08-226-486-1 ..

Align seg 1/1 to: US-08-226-486-1 from: 1 to: 1377

1 MetThrTyrProArgArgTyrArgArgArgHisArgProArg... 16
  ::::: ||| ::::: ||| ||| ||| :::::
311 TTGAGTATTGGCGCAAAAGGTCTCTCGAACGAGG...AGACCGAAGAC 357
  ::::: ||| ::::: ||| ||| ||| |||
17 .....SerHisLeuGlyGlnLeuArgA 25
  ::::: ||| ||| ||| |||
358 AGCAGGGGAAGAAAAGTGGAGTACGAGACTGACCCAGCGTTACTAA 407
  ::::: ||| ||| ||| |||
25 rgArgProTyrLeuValHisProArgHisArgTyrArgTyrArgGlys 41
  || ||| ||| ||| ||| ||| |||
408 GA.....TGTCGCTGCTCTGCTCATTTTCAGGATTTGGAGAGGACG 451
  ::::: ||| ||| ||| ||| |||
42 AsnGlyIlePheAsnThrArgLeuSerArgThrPheGlyTyrThrVally 58
  ::::: ||| ||| ||| ||| |||
452 GGG.....AGATTAAGT.....TTGAA 468
  ::::: ||| ||| ||| ||| |||
58 sArgThrThrVal.....ThrThrProSerTyr 68
  ::::: ||| ||| ||| ||| |||
469 GCATCGAGGTGCGTAGCGTGGCGATGCGCGAGTTCCTTCATCGT 517
  ::::: ||| ||| ||| ||| |||
68 laValAspMetMetArgPheLysleAspAsp..... 78
  ||| ::::: ||| ||| ||| |||
518 T....GACAGCGAGGAATACAGATTACGAACTAAAGTTTCACGGTATG 564
  ||| ::::: ||| ||| ||| |||
79 ...PheValProProGlyGlyGlyThrAsnLysIleSerIleProPheGln 94
  ||| ||| ||| ||| ||| ||| |||
565 AAGTTTGTCCCA.....GTTGCGGTTCCTCATGA 593
  ||| ||| ||| ||| ||| ||| |||
94 uTyrTyrArgIleArgLysValLysValGluPheTyrProCysSerProI 111
  ||| ||| ||| ||| ||| ||| |||

```

```

594 ATTTAAAGGGATTGCAAGAGAGAGAGTTTCACTTCGTGGAAGACTCCCGCG 643
111 leThrGlnGlyAspArgGlyValGlySerThrAlaValIleLeuAspAsp 127
  ::|||::: ::|||::: ::|||::: ::|||:::
644 TTACGGAGATACCAACGCGCTT.....AAGACAATGCTCACTAGGAT 687
  ::|||::: ::|||::: ::|||::: ::|||:::
128 AsnPheValThrLysAlaThrAlaLeuThrTyrAspProTyr...Valas 143
  ::|||::: ::|||::: ::|||::: ::|||:::
688 AGTTTTTCTGCACGTAAAGTAAAGCAGCATGGAGAGCCCGCAGCAGCTTGT 737
  ::|||::: ::|||::: ::|||::: ::|||:::
143 nTyrSerSerArgHisThrIleProGlnProPheSerTyrHisSerArgT 160
  ::|||::: |||::: |||::: |||::: |||:::
738 GGTACACACGGTGGGTACCGTC.....TACCACACCGCGTT 772
  ::|||::: |||::: |||::: |||::: |||:::
160 yr 160
773 TT 774

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-403-852D-14

seq_documentation_block:
: Sequence 14, Application US/08403852D
: Patent No. 5891695
: GENERAL INFORMATION:
: APPLICANT: Blanc, Veronique
: APPLICANT: Blanche, Francis
: APPLICANT: Crouzet, Joel
: APPLICANT: Jacques, Nathalie
: APPLICANT: Lacroix, Patricia
: APPLICANT: Thibaut, Denis
: APPLICANT: Zagorec, Monique
: APPLICANT: Debussche, Laurent
: APPLICANT: De Crecy-Lagard, Valerie
: TITLE OF INVENTION: Polypeptides Involved In The
: TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
: TITLE OF INVENTION: Coding For These Polypeptides And Their Use
: NUMBER OF SEQUENCES: 43
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
: STREET: 1300 I Street, N.W., Suite 700
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005-3315
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/403,852D
: FILING DATE: 10-MAY-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/FR 93/00923
: FILING DATE: 25-SEP-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: FR 92/11441
: FILING DATE: 25-SEP-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Meyers, Kenneth J.
: REGISTRATION NUMBER: 25,146
: REFERENCE/DOCKET NUMBER: 03806.0054-00000
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 408-4000
: TELEFAX: (202) 408-4400
: INFORMATION FOR SEQ ID NO: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 474 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: HYPOTHETICAL: NO

```



```
725 CCCTCGGCTTCAGAGACTTCACTATGCGCGGGGATCCCTTTCTC 774
140 rGlyrValAsnTyrSerSerArgHisThrIleProGlnProPheSerTyr 156
775 CTTTCTCGCTGCCCAAGAGCTGCTCACACGACGCGCTGTTTCATCC 824
157 HisSerArgTyrPheThrProLysProValLeuAspSerThrIleAspTyr 173
825 CAT.....ACCACATAGGCTTTGTAAAGCAACATCAATACATT 862
173 rPheGlnProAsnAsnLysArgAsnGlnLeuTyr 184
863 CTTCTCCCCAGCTAAGAGCAATAATAGATGG 896
```

seq\_name: /cgn2\_5/ptodata/1/ina/5C\_COMB.seq:US-08-910-973-12

```
seq_documentation_block:
; Sequence 12, Application US/08910973
; Patent No. 5795723
; GENERAL INFORMATION:
; APPLICANT: Tapscott, Stephen J.
; APPLICANT: Olson, James M.
; TITLE OF INVENTION: Expression of Neurogenic bHLH Genes in Primitive Neuroectoder
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC
; STREET: 1420 Fifth Avenue, Suite 2800
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/910, 973
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/239, 238
; FILING DATE: 06-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US95/05741
; FILING DATE: 08-MAY-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/17532
; FILING DATE: 30-October-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Sheiness, Diana K.
; REGISTRATION NUMBER: 35,356
; REFERENCE/DOCKET NUMBER: FPCR-1-10958
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-682-8100; 206-224-0735 (direct)
; TELEFAX: 206-225-0779
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; TYPE: nucleic acid
; LENGTH: 1268 base pairs
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: 20A1 (neuroD3)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 55..768
; US-08-910-973-12
```

```
alignment_scores:
Quality: 80.00 Length: 188
Ratio: 0.909 Gaps: 9
Percent Similarity: 46.809 Percent Identity: 25.000

alignment_block:
US-09-209-961-5 x US-08-910-973-12 ..
Align seg 1/1 to: US-08-910-973-12 from: 1 to: 1268

6 ArgArgTyrArgArgArgHisArgProArgSerHis.....LeuG1 20
265 AGCGCGCGCGCGCGCGCGGTCGCGGAGCGGGTCCGCTCGGAGCGCGTGTGCA 314
20 yGlnIleLeuArgArgArgProTyrIleuValHisProArgHisArgTyrA 37
315 CTCGCTGCGCAGGAGCGCGCGCTCAAGGCCAACGATCGCGAGCGCAACC 364
37 rgTyrArgArgLysAsnGlyIlePheAsnThrArgLeuSer.....Arg 51
365 GCATGCACAACTTGAACGCGCGCTGGACGCTGCGACGCTGCGTGTGCC 414
52 ThrPheGlyTyrThrValLysArgThrThrValThrProSerThrPal 68
415 TCGTTCCCGCAGCAGCACCAAGCTCACCAAAATCGAGACGCTGGCTTCGC 464
68 aValAspMetMet.....ArgPheLysIleAspAspPheV 80
465 CTACAACTACATCTGGGCTCTGGCGGAGACACTGGCGCTGGCGGATCAAG 514
80 alProProGlyGlyThrAsnLysIleSerIleProPheGluTyrTyr 96
515 GGCTCGCGGAGCGGTGCC.....534
97 ArgIleArgLysValLysValGluPheTrpProCys.....108
535 CGGAGCGGCTCTCGCGCGCGAGTGGTCTCCCTGCTCGCGGTCCCC 584
109 SerProIleThrGlnGlyAsp.....ArgGlyValGlySerThrA 122
585 AAGCGCGCGCAGCGAGCGGAGTCTGGGGCTCAGGTGCGCGCGCGCT 634
122 laValIleLeuAspAsnPheValThrLysAlaThrAlaLeuThrTyr 138
635 CCGCGCTCTCTGACCCAGTACGCCAGCGCGCTCCGAGAGACTTCACTAC 684
139 .....AspProTyrValAsnTyrSerSer.....Ar 147
685 CGCGCGCGGCGACCCGTGTTTCTCTCCCAAGCGCTGCCCAAGACTTGCT 734
147 gHisThrIleProGlnProPheSerTyrHis.....Sera 159
735 CCACACACAGCGCGCTGTTTCATTCTTACCACCTAGGCCCTTTGTAGACT 784
159 rgTyrPheThrPro 163
785 GTTACTTTCCCGCT 798

seq_name: /cgn2_5/ptodata/1/ina/5B_COMB.seq:US-08-314-309A-5

seq_documentation_block:
; Sequence 5, Application US/08314309A
; Patent No. 5677141
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: FUKAGAWA, MASAO
; APPLICANT: IWAMI, MORIYO
; APPLICANT: ARAMORI, ICHIRO
; APPLICANT: KOJO, HITOSHI
; TITLE OF INVENTION: PROCESS FOR PRODUCING 7-AMINOCEPHEM
; TITLE OF INVENTION: COMPOUND OR SALTS THEREOF
; NUMBER OF SEQUENCES: 34
```







OM of: US-09-209-961-5 to: EST:\* out\_format : pfs

Date: Dec 23, 1999 5:42 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:

-MODEL=framet.p2n.model -DEV=xlp  
-O=/cnp2.1/USPTO.spool/US09209961/runat.22121999.101043.17901/app\_query.fasta.1  
-DB=EST-QFMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500  
-GAPEXT=0.050 -XGAPOP=10.000 -YGAPEXT=0.500 -FGAPOP=6.000  
-DELEX=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEX=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdd  
-LIST=45 -DOCALLIGN=200 -THR SCORE=score -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pfs -NORM=stat -USER=US09209961 -NCPU=6 -ICPU=3 -NO\_XLPXY  
-WAIT -THREADS=1

Search information block:

Query: US-09-209-961-5

Query length: 233

Database: EST:\*

Database sequences: 4089388

Database length: 1713965092

Search time (sec): 1572.090000

score\_list:

Sequence	Strd Orig	zScore	Escore	Len	Documentation
gb_est32:AU070895	+	93.00	152.57	676	! AU070895 AU070895 Rice cDNA fro
gb_gss10:A0638908	+	86.00	147.84	316	! A0638908 927P1-14H7.TP 927P1.T
gb_est31:A1673888	-	88.50	145.35	630	! A1673888 605038F09.xl 605 - End
gb_est36:A1895135	+	87.50	144.08	600	! A1895135 EST264578 tomato callu
gb_est27:A1460429	+	85.00	144.06	387	! A1460429 sa81d12.yl Gm-cl004 G
gb_est32:A1736822	+	85.00	141.64	36.09	! A1736822 sb34a05.yl Gm-cl012 G
gb_gss8:A0577736	+	87.00	141.39	719	! A0577736 nbxb0091H09r CUGI Rice
gb_est37:A1946727	+	86.00	141.07	623	! A1946727 bs30a08.yl Drosophila
gb_est36:A1900445	+	82.50	141.04	38.98	! A1900445 sc05h05.yl Gm-cl012 G
gb_gss2:CMS017H1	-	88.00	140.75	915	! A1108048 Drosophila melanogaste
gb_est36:A1392143	+	84.00	139.76	45.93	! A1392143 NCC2D977 Confidial Neur
gb_est32:A069671	-	81.50	139.46	47.73	! A069671 AU069671 Rice callus C
gb_gss4:A0175288	-	83.00	138.93	51.14	! A0175288 HS_3210.AL.G12.MR CIT
gb_est27:A1441014	+	82.00	138.71	52.60	! A1441014 sa03g11.yl Gm-cl004 G
gb_est37:A1941495	+	82.00	138.60	53.30	! A1941495 sb90g09.yl Gm-cl017 G
gb_est32:A1735906	+	82.00	138.43	54.51	! A1735906 sb20d09.yl Gm-cl007 G
gb_est27:A1416813	+	82.00	138.38	54.86	! A1416813 sa18c05.yl Gm-cl005 G
gb_est28:A1495208	+	82.00	138.38	54.86	! A1495208 sb02a08.yl Gm-cl004 G
gb_est27:A1444041	+	82.00	138.35	55.03	! A1444041 sa31e11.yl Gm-cl004 G
gb_est28:A1460767	+	82.00	138.35	55.03	! A1460767 sa69a05.yl Gm-cl004 G
gb_est27:A1441174	+	81.50	138.24	55.81	! A1441174 sa52c03.yl Gm-cl004 G
gb_est28:A1495366	+	82.00	138.21	56.08	! A1495366 sa97b04.yl Gm-cl004 G
gb_est27:A1441122	+	82.00	138.13	56.61	! A1441122 sa59a12.yl Gm-cl004 G
gb_est32:A1736457	+	82.00	137.87	58.54	! A1736457 sb28h05.yl Gm-cl009 G
gb_est36:A1900949	+	82.00	137.75	59.43	! A1900949 sc17e10.yl Gm-cl013 G
gb_est28:A1461214	+	82.00	137.73	59.61	! A1461214 sa76f05.yl Gm-cl004 G
gb_est27:A1460639	+	82.00	137.71	59.79	! A1460639 sa71d09.yl Gm-cl004 G
gb_est27:A1416528	+	82.00	137.62	60.50	! A1416528 sa10c04.yl Gm-cl003 G
gb_est28:A1495104	+	82.00	137.50	61.39	! A1495104 sb03e02.yl Gm-cl004 G
gb_est37:A1941368	+	85.00	136.97	65.68	! A1941368 sc12c06.yl Gm-cl013 G
gb_est27:A1442364	+	82.00	136.95	65.89	! A1442364 sa26a06.yl Gm-cl004 G
gb_est36:A1900279	+	82.00	136.55	69.35	! A1900279 sa41h09.yl Gm-cl004 G
gb_est19:AA751860	-	82.00	136.47	70.09	! A1900279 sc03d06.yl Gm-cl012 G
gb_est36:A1900279	+	82.00	136.25	72.12	! A1900279 sc03d06.yl Gm-cl012 G
gb_gss10:A0641744	-	82.00	136.07	73.79	! A0641744 RPC193-DpnII-2616.TJ
gb_est36:A1885176	+	83.50	135.93	73.79	! A1885176 wa90f05.xl NCI-CGAP.B
gb_est32:A1727474	+	79.50	135.85	75.12	! D22673 R1C00842A Rice callus O
gb_est36:A1881562	-	81.50	134.95	85.11	! A1881562 BNLGH18050 Six-day Cot
gb_est28:A1494957	+	78.00	134.73	87.59	! A1494957 sa93b05.yl Gm-cl004 G
gb_est18:AA680493	+	81.50	134.56	89.53	! AA680493 SWAMCA3090SR Bruglia m
gb_est32:A0069559	+	80.00	134.40	91.38	! A0069559 AU069559 Rice callus d

gb\_est31:A1692946 + 81.00 134.21 93.63 513 ! A1692946 wd89f08.xl NCI\_CGAP  
gb\_est36:A1881449 - 81.50 134.16 94.22 563 ! A1881449 606069810.yl 606 -  
gb\_est5:N31093 - 80.50 134.01 96.00 479 ! N31093 yx63h03.r1 Soares mel

seq\_name: gb\_est32:AU070895

seq\_documentation\_block: 676 bp mRNA EST 10-JUN-1999  
LOCUS AU070895  
DEFINITION AU070895 Rice cDNA from young root Oryza sativa CDNA clone  
R10430.1A, mRNA sequence.

ACCESSION AU070895

NID 95038785

VERSION AU070895.1 GI:5038785

KEYWORDS EST.

SOURCE Oryza sativa.

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;  
Poaceae; Oryza.

REFERENCE 1 (bases 1 to 676)

AUTHORS Yamamoto K. and Sasaki, T.

TITLE Rice cDNA from young root

JOURNAL Unpublished (1999)

COMMENT On Jun 22, 1998 this sequence version replaced gi:3246692.

Contact: Takuji Sasaki

National Institute of Agrobiological Resources

Rice Genome Research Program

2-1-2 Kannondai, Tsukuba

Ibaraki,

Japan 305

Tel: 0298-38-7441

Fax: 0298-38-7468

Email: tsasaki@r.affrc.go.jp

PROJECT "RGP"

Location/Qualifiers

source

1. 676

/organism="Oryza sativa"

/strain="Nipponbare"

/db\_xref="taxon:4530"

/clone="R10430.1A"

/clone\_lib="Rice cDNA from young root"

/tissue\_type="young root"

BASE COUNT 130 a 225 c 211 g 105 t 5 others

ORIGIN

alignment\_scores:

Quality: 93.00 Length: 196

Ratio: 1.208 Gaps: 13

Percent Similarity: 39.286 Percent Identity: 26.531

alignment\_block:

US-09-209-961-5 x AU070895

Align seg 1/1 to: AU070895 from: 1 to: 676

4 ProArgArgArgTyrArgArgArg..... 11

|||||..... 13

90 CCTCGGCGGAGACACGCGCGGAGGACTACGTCACTGCACACAGCGCG 139

|||||..... 15

12 .....ArgHisArgPro..... 15

|||||..... 13

140 CGGCGGCGGAGCGCGCGGAGGAGTGGGAGCCCAAGTCGCCAGC 189

16 ..ArgSerHisLeu...GlyGlnIleLeuArgArgProTrpLeuVal 30

|||||..... 13

190 TTCGCGAGAGTACGCGCGGAGCGCGGAGTCCGCGCTGNCAGC 239

|||||..... 35

31 HisProArgHisArg..... 35

|||||..... 289

240 ACTCGGCGGCGCGGAGTACGCGGAGAGAACATCTCTGGGGCTCGCGGGCGGC 289

```

36 .....TyrArgTyrArgArgLysAsnGlyIlePheAsnt 47
   |||||
290 GCCTGGAGCGCGCGACGCGTGGCGTGT..... 320
47 hrArgLeuSerArgThrPheGlyTyrThrValLysArgThrThrValThr 63
   |||||
321 .....GGGTGGCGGAGAGAGAACTACCACTAGC 350
64 ThrProSerTrpAlaValAspMetMetArgPheLysIleAspPheVa 80
   |||||
351 ACACCAACAGCTGGAC.....CGGGCAAGGTGTGGCGCACTT 391
80 lProGlyGly.....GlyThrAsnLysIleSerIleProp 93
   |||||
392 ACACCAAGTGTGGCGCAAGTGGTGGCATCGGTG..... 431
93 heGluTyrTyrArgIleArgLysValLysValGluPheTrpProCysSer 109
   |||||
432 .....CGCCGCGGTGTGGCG..... 449
110 ProIleThrGlnGlyAspArgGlyValGlySerThrAlaValIleLeuA 126
   |||||
450 .....GGCAACCGCGCGGTG..... 465
126 spAspAspPheValThrLysAlaThrAlaLeuThrTyrAspProTyrVal 142
   |||||
466 .....TTCATCACC.....TGCACTAGACCCCGCGGC 495
143 AsnTyrSerSerArgHisThrIleProGlnProPhe 154
   |||||
496 AACTTCAACGCGGAG.....CGCCGCTTC 519

```

seq\_name: gb\_gss10:AQ638908

```

seq_documentation_block:
LOCUS AQ638908 316 bp DNA GSS 08-JUL-1999
DEFINITION 927P1-14H7.TP 927P1 Trypanosoma brucei genomic clone 927P1-14H7,
genomic survey sequence.
ACCESSION AQ638908
NID 95115618
VERSION AQ638908.1 GI:5115618
KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei
Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
REFERENCE 1 (bases 1 to 316)
AUTHORS El-Sayed,N., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
Gerrard,C., Leach,V., de Jong,P., Ullu,E., Melville,S.,
Donelson,J., Fraser,C. and Adams,M.
Determination of clone end sequences from Trypanosoma brucei TREU
927/4 P1 library
JOURNAL Unpublished (1999)
COMMENT Other GSSs: 927P1-14H7.TV
Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@tigr.org
For clone/filter availability, please contact Sara Melville
(sm160@ole.bio.cam.ac.uk). P1 end sequences search page:
http://www.tigr.org/tdb/mdb/tbdb/.
Seq primer: SP6
Class: P1 ends.

```

FEATURES  
source

Location/Qualifiers  
1..316  
/organism="Trypanosoma brucei"  
/strain="TREU927/4"  
/db\_xref="taxon:5691"  
/clone="927P1-14H7"  
/clone\_lib="927P1"

/note="Vector: pAD10sacBII; Site\_1: Bam HI; Constructed by Sara Melville, University of Cambridge, UK and Nancy Shepherd, Dupont Merck, Wilmington, DE. Genomic DNA was isolated from Trypanosoma brucei (stock TREU927/4) and partially digested with Sau 3AI. DNA fragments were cloned into the Bam HI site of pAD10sacBII vector (Genbank accession U09128). The average insert size is 65 Kb. Coverage: approx 4.4 x the haploid non-minichromosomal genome."

BASE COUNT 90 a 74 c 69 g 82 t 1 others  
ORIGIN

alignment\_scores:  
Quality: 86.00 Length: 49  
Ratio: 2.867 Gaps: 1  
Percent Similarity: 61.224 Percent Identity: 38.776

alignment\_block:

US-09-209-961-5 x AQ638908/rev ..

Align seg 1/1 to reverse of: AQ638908 from: 1 to: 316

```

65 ProSerTrpAlaValAspMetMetArgPheLysIle..... 76
|||||
305 CCTTCGTGGCCCTTTCGACGATCTCTTCGAGTCCGAGCAACTCAAGA 256
77 ....AspAspPheValProGlyGlyThrAsnLysIleSerIleP 92
|||
255 TTAGGATGATGAGGAGTCCCGGAGGAGGATCGAACCCTGGACAATCG 206
92 ropheGluTyrTyrArgIleArgLysValLysValGluPheTrpPro 107
|||||
205 AGTTCAACTTCACGCGGATCATGAAGATAGCTTTCTTAGACTACCCCT 159

```

seq\_name: gb\_est31:AI673888

```

seq_documentation_block:
LOCUS AI673888 630 bp mRNA EST 19-MAY-1999
DEFINITION 605038f09.x1 605 - Endosperm cDNA library from Schmidt lab Zea mays
CDNA, mRNA sequence.
ACCESSION AI673888
NID 94874368
VERSION AI673888.1 GI:4874368
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
Poaceae; Zea.
REFERENCE 1 (bases 1 to 630)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
JOURNAL Unpublished (1999)
COMMENT On Mar 10, 1998 this sequence version replaced gi:2948185.

```

Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Plate: 605038 row: F column: 09.

FEATURES  
source

Location/Qualifiers  
1..630  
/organism="Zea mays"  
/cultivar="Ohio43"  
/db\_xref="taxon:4577"  
/clone\_lib="605 - Endosperm cDNA library from Schmidt lab"  
/tissue\_type="pericarp, embryo, and endosperm"  
/dev\_stage="10 days post-pollination"

/lab\_host="DH5(alpha)"  
 /note="Organ: Kernel; Vector: PAD-GAL4-2"; Site\_1: EcoRI;  
 Site\_2: XhoI; Kernel endosperm cDNA library from Schmidt  
 lab"

BASE COUNT 131 a 185 c 239 g 74 t 1 others  
 ORIGIN

alignment\_scores:  
 Quality: 88.50 Length: 135  
 Ratio: 1.609 Gaps: 6  
 Percent Similarity: 40.741 Percent Identity: 25.926

alignment\_block:  
 US-09-209-961-5 x AI673888/rev ..  
 Align seg 1/1 to reverse of: AI673888 from: 1 to: 530

```

1 MetThrTyrProArgArgArgTyrArgArgArgHisArgProArgse 17
...          |||||          |||||          |||||
541 CPACCTTCCCTCCGCGG...AGCGGAGGAGCGGCGCCGTCA 495
17 rHisLeuGlyGln.....lleuArgArgArgProTrp. 28
...          |||||          |||||          |||||
494 TCTACTCGTCCCGCCACCTCCTGCTGCTTCGCGCGCGCGCCTTGT 445
29 .....LeuValHisProArgHisArgTyrArgTrp 38
...          |||||          |||||          |||||
444 GCGGAGGCTTACTACCGCGCCACCGCGGAGCGCGCGCATGG 395
39 ArgArgLysAsnGlyIlePheAsnThrArgLeuSerArgThrPheGlyTy 55
|||
394 CGG..... 392
55 rThrValLysArgThrThrValThrThrProSerTrpAlaValAspMetM 72
|||
391 .....TGGCGCGTCCGACG 376
72 etArgPheLysIleAspPheValProGlyGlyGlyThrAsnLy 88
|||
375 GCGGTGTGTGTACCGCGCGCGCGCGCGCGCCCAACCGTTCCT 326
88 sileSerileProPheGluTyrTyrArgIleArgLysValGluP 105
...|||
325 GCCCTACTCCCTCTACTCTACAGC..... 298
105 heTrpProCysSerProIleThrGlnIleAspArgGlyValGlySerThr 121
|||
297 .....CCGCGCGCGCGTGTGACCTCGCGCGCGCGCGCGCCGCCACC 253
122 Ala 122
...
252 TCG 250
```

seq\_name: gb\_est36:AI895135

seq\_documentation\_block:  
 LOCUS AI895135 600 bp mRNA EST 27-JUL-1999  
 DEFINITION EST264578 tomato callus, TAMU Lycopersicon esculentum cDNA clone  
 CLEC6L6, mRNA sequence.

ACCESSION AI895135

NID 95601037

VERSION AI895135.1 GI:5601037

KEYWORDS EST.

SOURCE tomato.

ORGANISM Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core  
 eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum;  
 Potatoe; Lycopersicon.

1 (bases 1 to 600)

REFERENCE Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T.,

Holt,I.E., Liang,F., Upton,J., Ronning,C.M., Craven,M.B.,

Fujii,C.Y., Bowman,C.L., Niernan,W., Fraser,C.M., Venter,J.C.,  
 Martin,G.B., Tanksley,S.D. and Giovannoni,J.  
 Generation of ESTs from tomato callus tissue  
 Unpublished (1999)  
 On Dec 20, 1995 this sequence version replaced gi:1136036.

TITLE  
 JOURNAL  
 COMMENT

Contact: David Frisch  
 Clemson University Genomics Institute  
 Clemson University  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Tel: 864 656 4366  
 Fax: 864 656 4293  
 Email: dfrisch@CLEMSON.EDU  
 5 prime sequence.

FEATURES

Location/Qualifiers

1..600

/organism="Lycopersicon esculentum"

/cultivar="TA496"

/db\_xref="taxon:4081"

/map="15q26.1: 8p21.3-p22"

/clone="CLEC6L6"

/clone\_lib="tomato callus, TAMU"

/tissue\_type="callus"

/dev\_stage="combined undifferentiated and shooting callus"

/lab\_host="XLI-Blue MRF"

/note="Vector: pBlueScript SK(-); Site\_1: EcoRI; Site\_2:

XhoI; CLEC - Tomato Callus EST Library. Directionally

cloned cDNAs inserted into pBlueScript SK(-) at 5' end

with EcoRI and 3' end with XhoI site."

BASE COUNT 191 a 134 c 105 g 169 t 1 others

ORIGIN

alignment\_scores:

Quality: 87.50 Length: 223

Ratio: 0.911 Gaps: 11

Percent Similarity: 43.049 Percent Identity: 21.525

alignment\_block:

US-09-209-961-5 x AI895135 ..

Align seg 1/1 to: AI895135 from: 1 to: 600

```

42 AsnGlyIlePheAsnThrArgLeuSerArgThrPheGlyTyrThrVally 58
|||||          |||          |||          |||
4 AATGGTGCCTTATTATACGTCACGCAATAAACTAATATCTTTCC 53
58 sArgThrThrValThrThrPro.....SerTrpAlaValAspM 71
|||          |||          |||          |||
54 TAGACCTACTGCTGAATTCCTCATATTAGGGGAGTGGGAAAGTG 103
71 etMetArgPheLysIleAspAspPheValProGlyGlyGlyThrAsn 87
.....          |||          |||          |||
104 ATGTACAAGCAGTTGTGAATCAGTTCTTACAAGTGGCAGCATGCAAT 153
88 LysIleSerIleProPheGluTyrTyrArgIleArgLysValGlyValG 104
|||          |||          |||          |||
154 AAA.....TCTGATGCATTTCTTATAAATGTCACACGAGGAGA 191
104 uPheTrpProCysSerProIleThrGlnGlyAspArgGlyValGlySerT 121
.....          |||          |||          |||
192 TTTATATCTTCTCTCA..... 207
121 hrAlaValIleLeuAspAsnPheValThrLysAlaThrAlaLeuThr 137
.....          |||          |||          |||
208 .....ACAATGCACATTC.....AAACTAAGC 231
138 TyrAspProTyrValAsnTyrSerSerArg..... 147
...          |||          |||          |||
232 GTGGAGAAAAAATAGACATATTCTTCATGTTAAGCCGCGTAATGAA 281
148 .....HisThrIlePro..... 151
|||          |||          |||          |||
```

```

282 TAATCTCATGTTTTCTCCGTTGCAATCAACCACTCACCGTCGTAGGAA 331
152 .....GlnProPheSerTyrHisSerArgTyrPheThr 162
332 CTGAGGGTGGCTACGTTAAACGGTTCAG.....TCAATATACATACA 375
163 ProLysProValLeuAspSerThrIleAspTyrPheGlnProAsnAsnLys 179
376 ATTTCCTCCG.....GGTCAACTATTGAT..... 399
179 sArgAsnGlnLeuTrpLeuArgLeuGlnThrSerGlyAsnValAspHisV 196
400 .....GTTTACTCGAAGCAAAATCAAAACCTCAACCACT 433
196 alGlyLeuGlyThrAlaPheGluAsnSerLysTyrAspGlnAspTyr... 211
434 ATTATATGGCGCGCAAGCCTACATAGCGTAGTCGGAGTTGATTTAAC 483
212 AsnIleArgValThrMetTyrValGlnPheArg...GluPheAsnLeuLys 227
484 AACATAACCAACCGCCGATCGTTGAGTTCGTGGAAATTAATCACTTC 533
227 sAspProProLeuGluPro 233
534 ATCACCAACCACTCCTTC 552

```

seq\_name: gb\_est27:AI460429

seq\_documentation\_block: 387 bp mRNA EST 09-MAR-1999  
 LOCUS AI460429 sa81d12.v1 Gm-cl004 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:  
 DEFINITION Gm-cl004-5712 5' similar to TR:Q39882 Q39882 NODULIN-26. ;, mRNA  
 sequence.

ACCESSION AI460429  
 NID 94313310  
 VERSION AI460429.1 GI:4313310  
 KEYWORDS EST.  
 SOURCE soybean.

#### ORGANISM

Glycine max  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core  
 eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;  
 Glycine.

1 (bases 1 to 387)

Shoemaker, R., Briner, D., Marra, M., Hillier, L., Kucaba, T.,  
 Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M.,  
 Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T.,  
 Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S.,  
 Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and  
 Wilson, R.

#### TITLE

Public Soybean EST Project  
 Unpublished (1999)  
 On Jun 5, 1998 this sequence version replaced gi:3189388.

#### COMMENT

Contact: Shoemaker R/Public Soybean EST Project  
 Public Soybean EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available through: Genome Systems, Inc. 4633 World  
 Parkway Circle St. Louis, Missouri 63134 For further information  
 call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)  
 427-3324 or contact: clones@genomesystems.com or  
 info@genomesystems.com web site: www.genomesystems.com  
 Possible reversed clone: similarity on wrong strand  
 Seq primer: -40RP from Gibco  
 High quality sequence stop: 280.

#### FEATURES

source  
 1. 387  
 Location/Qualifiers  
 /organism="Glycine max"  
 /db\_xref="taxon:3847"

```

/clone="GENOME SYSTEMS CLONE ID: Gm-cl004-5712"
/clone_lib="Gm-cl004"
/tissue_type="root"
/lab_host="XL10-Gold"
/note="vector: pBluescript II XR; Site_1: EcoRI; Site_2:
XhoI; Root cDNA. The mRNA was isolated from entire roots
of 8 day old 'Williams' seedlings which were propagated on
paper towels with distilled water. Stratagene's cDNA
Synthesis Kit (catalog #200401) was used to synthesize the
cDNA. First-strand synthesis was performed with 5-methyl
dCTP, hence the ligated cDNA is hemimethylated.
Stratagene's first-strand synthesis primer was used
[IGAGAGAGAGAGAGAGAGACTAGTCTCGAG(T)-18]. After
second-strand synthesis, the cDNA ends were 'polished'
with clone Pfu DNA polymerase, ligated to EcoRI adapters,
and phosphorylated. The XhoI site within the first-strand
synthesis primer was restricted by digestion with XhoI;
all XhoI sites in the cDNA would be protected by their
hemimethylated status. The cDNA constructs were
size-fractionated with a 500bp cutoff, using GibcoBRL Life
Technologies' cDNA Size Fractionation column. The column
eluent was then ligated into Stratagene's pBluescript II
XR Predigested vector (pBluescript II SK(+)) that had been
digested with EcoRI and XhoI, and phosphorylated. Both
the white and blue colonies appear to contain recombinant
plasmids with cDNA inserts. Blue colonies 9n-15) have been
sequenced, and possess putative cDNA inserts. This library
was constructed by Dr. Paul Keim & Virginia H. Coryell,
Department of Biology, Box 5640, Northern Arizona
University, Flagstaff, AZ 86011, Phone: 520-523-1078 (Dr.
Paul Keim), 520-523-1372 (Virginia H. Coryell), Fax:
520-523-7500, email: paul.keim@na.u.edu,
virginia.coryell@na.u.edu"

```

BASE COUNT 54 a 139 c 92 g 101 t 1 others  
 ORIGIN

alignment\_scores:  
 Quality: 85.00 Length: 79  
 Ratio: 2.024 Gaps: 2  
 Percent Similarity: 53.165 Percent Identity: 31.646

#### alignment\_block:

```

US-09-209-961-5 x AI460429 ..
Align seg 1/1 to: AI460429 from: 1 to: 387

5 ArgArgArgTyrArgArgArgHis.....ArgProAr 16
|||||  |||||||  |||||||  |||  |||  |||  |||  |||  |||
57 CGTCGCGCTCTCGTCGGCGCAACATCTCCGGCGCCACGTCACCCCG 106
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
16 gSerHisLeuGlyGlnIleLeuArgArgArgProTyrLeuValHisProA 33
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
107 CCCTCACCTTCGGCGC...CTTCGTCGGCGCAACATCACCTTCCTCCGC 153
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
33 rGHisArgTyrArgTyrArgArgLysAsnGlyIlePheAsnThrArgLeu 49
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
154 GGCATGCTCTACCTACATCATCGCCAGCTCCTCGGTCCATCGTCGCCCTCCCT 203
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
50 SerArgThrPheGlyTyrThrValLysArgTyrThrValThrThrProSe 66
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
204 CCTCTTGGCTTCGTACCGCTCCACTGTTCAGCATTCGGACTCTCCG 253
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
66 rTPAlaValAspMetMetArgPheLysIleAspAsp 78
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
254 CTGAGTGTGGAGTTGGCAACGCTTGTTGGTGTGGAGAT 290
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

seq\_name: gb\_est32:AI736822

seq\_documentation\_block: 493 bp mRNA EST 14-JUN-1999  
 LOCUS AI736822 sb34a05.v1 Gm-cl012 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:  
 DEFINITION Gm-cl012-297 5' similar to TR:Q25252 Q25252 TONOPLAST INTRINSIC



and genome sequencing of rice, we have constructed a BAC library from *Oryza sativa*, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 Kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."

BASE COUNT 108 a 234 c 246 g 131 t  
ORIGIN

alignment\_scores:  
Quality: 87.00 Length: 135  
Ratio: 1.261 Gaps: 7  
Percent Similarity: 51.111 Percent Identity: 28.889

alignment\_block:

US-09-209-961-5 x AQ577736 ..

Align seg 1/1 to: AQ577736 from: 1 to: 719

```

6 ArgArgTyrArgArgArgArgHisArgProArgSerHisLeuGlyGlnI 22
|||||
171 CGCGCGGGTGGCGCGCGAGTTCGT.....CACCTCGGSCACCT 211
|||||
22 eLeuArgArgArgProTrpLeuValHisPro..... 32
|||||
212 TCCTCATCGACGGCGCAACATCGAACATCCAGTCTACCCCGACAGGTGG 261
|||||
33 .....ArghisArgTyr..... 36
|||||
262 AAGCAGAGATGACGGCGGTAGTGTCTCTCTCTCTCTCTCTCTCTCTCGG 311
|||||
37 ArgTrpArgArgLysAsnGlyIlePheAsnThrArgLeuSerArgThrPh 53
|||||
312 CGGAGCAACCGCGTGGCGAAGTACACGCTGAGCCCTTCGGAGAACGG 361
|||||
53 eGlyTyrThrValLysArgThr.....ThrValThrThrP 65
|||||
362 CGCGAGTCTCGTACACGGAGCTGACGATCGGTTCGACACACGGTGGCG 411
|||||
65 roSerTrpAlaValasp.....MetMetArgPheLysIleAspPhe 79
|||||
412 CTTCTCGGGATTCCTCCAGGTTCTATGGAGAGCCCGCGCTCGGCAATGG 461
|||||
80 Val.....ProProGlyGlyGlyThrAsnLysIleSerIleProPh 93
|||||
462 CTGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 511
|||||
93 eGlyTyrTyrArgLysValLysValLysValGlu.PheTrpProCysSer 109
|||||
512 CTCCTGACCGCTCATAGGGAGCCCTCGACCGAGGGCGTGGCGCGCGTGG 561
|||||
110 Pro 110
|||
562 CCG 564

```

seq\_name: gb\_est37:AI946727

```

seq_documentation_block: 623 bp mRNA EST 17-AUG-1999
LOCUS AI946727
DEFINITION bs30a08.y1 Drosophila melanogaster adult testis library Drosophila
melanogaster cDNA clone bs30a08 5', mRNA sequence.
ACCESSION AI946727
NID 95737153
VERSION AI946727.1 GI:5737153
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera.
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

```

# REFERENCE

1 (bases 1 to 623)  
Andrews, J., Bouffard, G. and Oliver, B.  
Drosophila melanogaster testis expressed sequence tags  
Unpublished (1999)  
On Jun 5, 1998 this sequence version replaced gi:3189407.

Contact: Brian Oliver  
Laboratory of Cellular and Developmental Biology  
NIDDK, National Institutes of Health  
6 Center Drive MSC 2715, Bldg 6, Rm B1-13, Bethesda, MD 20892 USA  
Email: oliverhelix.nih.gov,  
Fax: (301) 496 5239

http://www.niddk.nih.gov/intram/people/boliver.htm  
Tissue isolation and library construction performed at the National  
Institute of Diabetes and Digestive and Kidney Diseases, NIH (see  
http://www.niddk.nih.gov/intram/people/boliver.htm). DNA sequencing  
and analyses performed by National Institutes of Health Intramural  
Sequencing Center (NISC; see http://www.nisc.nih.gov). Chromatogram  
data were analyzed and evaluated for high quality using the ted  
program (Gleeson T and Hillier L, 1991).  
Plate: 30 row: a column: 08  
Seq primer: M13Rpl reverse primer (ABI).

# FEATURES

source

```

1..623
Location/Qualifiers
/organism="Drosophila melanogaster"
/strain="y[*] w[67cl]/Y"
/db_xref="taxon:7227"
/map="703C03; 1"
/clone="bs30a08"
/clone_lib="Drosophila melanogaster adult testis library"
/sex="male"
/dev_stage="1-5 day adult"
/lab_host="SOLR (Stratagene)"
/note="organ: testis; Vector: pBlueScript SK (Stratagene);
Site_1: EcoR I; Site_2: Xho I; Testes dissected from 1-5
day adult y[*] w[67cl]/Y males raised at 25°C. RNA
isolated using Trizol (Life Technologies) and a single
round of Poly(A)+ selection using Oligotex (Qiagen). cDNA
library constructed using Stratagene ZAP-cDNA synthesis
kit. Oligo dt-primed, size fractionated -1-6 kb, and
directionally cloned at EcoRI and XhoI in Uni-ZAP XR.
Following a single round of amplification pBlueScript SK
phagemids were mass excised. A distribution channel for
clones is being sought, but not currently available.
Requests for clones cannot be honored."
BASE COUNT 189 a 141 c 184 g 109 t
ORIGIN

```

alignment\_scores:

Quality: 86.00 Length: 135  
Ratio: 1.211 Gaps: 6  
Percent Similarity: 52.593 Percent Identity: 29.630

alignment\_block:

US-09-209-961-5 x AI946727 ..

Align seg 1/1 to: AI946727 from: 1 to: 623

```

4 ProArgArgTyrArgArgArgHisArgProArgSer...HisLe 19
|||||
259 CGCGGAAATCGCCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 308
|||||
19 uGlyGlnIleLeuArgArgProTrpLeuValHisProArgHisArg 36
|||||
309 AACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 346
|||||
36 yrArgTrpArgArgLysAsnGlyIlePheAsnThrArgLeuSerArgThr 52
|||||
347 GGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 393
|||||
53 PheGlyTyrThrValLysArgThrValThrProSerTrpAlaVa 69
|||||

```



collaboration with the European Drosophila Genome Project (EDGP) - <http://www.edgp.ebi.ac.uk> -. This *Drosophila melanogaster* BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelOBAC11.

```

FEATURES              Location/Qualifiers
    source            1..915
                        /organism="Drosophila melanogaster"
                        /plasmid="pBeloBAC11"
                        /db_xref="taxon:7227"
                        /clone_lib="DrosBAC"
                        /clone="BACN171Q3"
                        /notes="end : T7"
BASE COUNT           198 a   233 c   188 g   281 t   15 others
ORIGIN
alignment_scores:
    Quality:          88.00
    Ratio:             1.239
    Percent Similarity: 52.593
                        Gaps: 6
                        Percent Identity: 29.630
alignment_block:
US-09-2909-961-5 x CNS017HI/rev ..

```

```

Align seg 1/1 to reverse of: CNS017HI from: 1 to: 915

4  ProArgArgArgTyrArgArgArgArgHisArgProArgSer...HisLe 19
|||||:||||| |||||||:|||||:|||||:|||||:|||||:|||||:
406  CCGCAATCGCCCAAGGAGAGGAGGAGGAGGCCAGGAGCGCGAGAAG 357

19  uGlnGlnLeuArgArgArgProTrpLeuValHisProArgHisArgT 36
:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
356  AAGCAGAAGGAAAAAGAGAAGGCCG.....AAAAGGAGAAGGC 319

36  yrArgTrpArgArgLysAsnGlyIlePheAsnThrArgLeuSerArgThr 52
||| |||||:|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|
318  GGAGAAGGAGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG... 272

53  PheGlyTyrThrValLysArgThrThrValThrProSerTrpAlaVal 69
:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
271  .....AGAGGAAAAAGAGGAGGAGGCCCAAGGGCAGA 237

69  lAspMetMetArgPhePheLysIleAspSerPheValProGlyGlyGlyT 86
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:
236  GAGAAGGGCAAGGAGGGCAAGTAAAGCACTATAGACCCCAAGGACATG 187

86  hrAsnLysIleSerIle...ProPheGluTyrTyrArgIleArgLysVal 101
:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
186  TCAAGGAGCAGGATGTTTCCGCCCAAAATCTCATGAATCCATATCGTT 137

102  .....LysValGluPheTrpProCysSerProIleIleTh 112
:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
136  TCTAAACKTCAAAATTTCAAGTGCATTATATGTCATACCTAACGATAAC 87

112  rGlnGlyAspArgGlyValGlySerThrAlaValIleLeuAspAspAsn 129
| ||||:| |||||:| |||||:| |||||:| |||||:| |||||:|
86  TATTGTCAG.....TCCATTGCCGCGTGTGATGGAGAGCKTCT 49

129  heVal 130
||:|:|
48  TTATT 44

```

seq\_name: qb\_est26:AI392143

seq_documentation_block:			
LOCUS	AI392143	499 bp	mRNA
DEFINITION	NC02D9T7	Conidial <i>Neurospora crassa</i>	EST 03-FEB-1999
		triosephosphate isomerase, mRNA sequence.	
ACCESSION	AI392143		
NID	q4219950		

VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	COMMENT	
AI32143.1	GI:4219950	EST.	Neurospora crassa.	Neurospora crassa.	Neurospora crassa.	Neurospora crassa	Neurospora crassa	Fungal Genet. Biol. 21	97435549	On Jan 14, 1998 this s
			Eukaryota; Fungi; Ascomycota; Neurospora crassa	Sordariales; Sordariaceae	1 (bases 1 to 499)	Nelson, M.A., Kang, S., Mitchell, P.M., Cushing, T., Errett, A., Miller, R., Ortega, J., Trujillo, R., Valentine and Navig, D.O.	Expressed sequences from Neurospora crassa			

Contact: Natvig, D.O./Nelson, M.A.

Department of Biology  
University of New Mexico  
Casteretter Hall, Albuquerque, NM 87131, USA  
Tel: 505 277 3411  
Fax: 505 277 0304  
Email: [nqpebiology.unm.edu](mailto:nqpebiology.unm.edu).

FEATURES  
source  
1. .499  
Location/Qualifiers  
Lunar: hypetrology:am:cua.

```

/organism="Neurospora crassa
/strain="74-OR23-1V A"
/db_xref="taxon:5141"
/map="21"
/clone="NC2D9"
/clone_lib="Conidial"
/sex="Mating type A"
/tissue_type="Conidia"
/dev_stage="germinating conid
/note="mRNA isolated from ger
vogel's, 2% sucrose for 4.5 h
cloned into pBluescript SK(-)
system (Stratagene, La Jolla,
88 a 137 c 146 q 128 t
BASE COUNT

```

```

alignment_scores:      84.00      Length: 93
                      Ratio: 1.714      Gaps: 5
Percent Similarity: 52.688      Percent Identity: 31.183

alignment_block:
US-09-209-961-5 x AI392143/rev ..

Align seg 1/1 to reverse of: AI392143 from: 1 to: 499

6 ArgArgTyrArgArgArg.....HisArgProArgSe 17
|||||
367 CGTCGTATCGGTACGAGCCCATCTGGGCCATTGGCACC CGCCCTCGTCG 318

17 HisLeuGlyGln.....IleLeuArgArgProTrpLeuValH 31
|||||
317 CCACCAAGAGCAGGCCCGCAGGAGCTGCACCGCCATCCCGAGCTGGTCC 268

31 isProArgHisArgTyrArgTrpArgArgLysAsnGlyIlePheasnTrp 47
:::
267 AGCAGAA.....CGTCAGCGCACAAGGTTGCTGAGGGACC 233

48 ArgLeuSerArgThrPheGlyTyrThrValLysArgThr..... 60
|||||
232 CGCATCTCT...ACGGTGCTCTGTCACCGCCAGAACTGCAAGGACCTT 186

61 .....ThrValThrThrProSerTrpAlaValAspMetMeta 73
|||||

```



371 GATGAGGACCAGGGCAGGGCTTGGCAGCGGGGCTCTGGGCCGTTCTGTCCA 322



Public Soybean EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.edu  
 This clone is available through: Genome Systems, Inc. 4633 World  
 Parkway Circle St. Louis, Missouri 63134 For further information  
 call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)  
 427-3324 or contact: clones@genomesystems.com or  
 info@genomesystems.com web site: www.genomesystems.com  
 High quality sequence stop: 371.

## FEATURES

## Source

```

1. .394
   Location/Qualifiers
     /organism="Glycine max"
     /db_xref="taxon:3847"
     /clone="GENOME SYSTEMS CLONE ID: Gm-cl017-761"
     /clone_lib="Gm-cl017"
     /tissue_type="vegetable buds of field grown plants"
     /lab_host="XL10-Gold"
     /note="Vector: pBluescript II XR; Site1: EcoRI; Site2:
XhoI; This cDNA library was constructed from mRNA isolated
from vegetable buds of field grown plants. The cDNA
library was prepared using the Stratagene pBluescript II
XR library construction kit. Complementary DNA was
synthesized from mRNA using a primer consisting of a poly
(dT) sequence with a XhoI restriction site. EcoRI
adapters were ligated to the blunt-ended cDNA fragments
followed by XhoI digestion. The cDNA fragments were
directionally cloned into the EcoRI-XhoI restriction site
of the pBluescript vector. The ligated cDNA fragments
were transformed into XL10-Gold host cells."
BASE COUNT      66 a 156 c  89 g  83 t
ORIGIN

```

## alignment\_scores:

```

Quality:      82.00      Length:      40
Ratio:        3.280      Gaps:        2
Percent Similarity: 62.500      Percent Identity: 52.500

```

## alignment\_block:

```
US-09-209-961-5 x AI941495 ..
```

```
Align seg 1/1 to: AI941495 from: 1 to: 394
```

```

5  ArgArgArgTyrArgArgArgHis.....ArgProAr 16
|||||||  |||||||:|||||
258 CGTCGCGTCTCCGTCGCGCCACATCTCCGCGCGCCACGTCACCCCG 307
|||||||  |||||||:|||||

16 gSerHisLeuGlyGlnIleLeuArgArgProTrpLeuValHisProA 33
|  |||||  ::  |||||||||
308 CCGTCACGTCGCGGC...CTTCGTCGCGGCACATCACCTCTCTCCGC 354

33 rgHisArgTyrArgTrpArg 39
|||||||  |||  |||
355 GGCATCGCTTACGTCATCGC 374

```

OM of: US-09-209-961-7 to: GenEmbl.\* out\_format : pfs  
Date: Dec 27, 1999 2:13 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:  
-MODEL=frame1.p2n.model -DEV=xlp  
-O=/cgn2\_1/USPTO\_spool/US09209961/runat\_22121999\_101043\_17910/app\_query.fasta.1  
-DB=GenEmbl -QFMR=fastap -SUFFIX=rg -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -OGAPOP=4.500  
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=escore -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pfs -NORM=stat -USER=US09209961 -NCPU=6 -ICPU=3 -NO\_XLPXY  
-WAIT -THREADS=1

## Search information block:

Query: US-09-209-961-7  
Query length: 59  
Database: GenEmbl.\*  
Database sequences: 780561  
Database length: 2137953050  
Search time (sec): 1994.760000

## score\_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
gb_vi:AF085695	- 311.00	679.19	7.9e-30	1768	AF085695 Porcine circovirus st
gb_vi:AF086834	- 311.00	679.19	7.9e-30	1768	AF086834 Porcine circovirus st
gb_vi:AF086835	- 311.00	679.19	7.9e-30	1768	AF086835 Porcine circovirus st
gb_vi:AF118095	- 311.00	679.19	7.9e-30	1768	AF086836 Porcine circovirus st
gb_vi:AF055393	- 295.00	652.69	2.4e-28	597	AF118095 Porcine circovirus ty
gb_vi:AF055394	- 295.00	643.13	8.0e-28	1767	AF055393 Porcine circovirus ty
gb_vi:AF055394	- 295.00	643.13	8.0e-28	1767	AF055394 Porcine circovirus ty
gb_vi:AF027217	- 295.00	643.13	8.0e-28	1768	AF027217 Porcine circovirus st
gb_vi:AF055391	- 295.00	643.13	8.0e-28	1768	AF055391 Porcine circovirus ty
gb_vi:AF055392	- 295.00	643.13	8.0e-28	1768	AF055392 Porcine circovirus ty
gb_vi:AF109399	- 295.00	643.13	8.0e-28	1768	AF109397 Bovine circovirus, cc
gb_vi:AF109399	- 295.00	643.13	8.0e-28	1768	AF109399 Porcine circovirus ty
gb_vi:AF147751	- 295.00	643.13	8.0e-28	1768	AF147751 Porcine circovirus ty
gb_vi:AF154679	- 295.00	643.13	8.0e-28	1768	AF154679 Porcine circovirus, c
gb_vi:AF156528	- 295.00	643.13	8.0e-28	1768	AF156528 Porcine circovirus, c
gb_vi:PCAG3185	- 295.00	643.13	8.0e-28	1768	AF223185 Porcine circovirus DN
gb_vi:AF109398	- 295.00	634.11	2.6e-27	1768	AF109398 Porcine circovirus ty
gb_vi:AF117753	- 286.00	622.84	1.1e-26	1768	AF117753 Porcine circovirus ty
gb_vi:AF112862	- 284.00	618.34	1.9e-26	1768	AF112862 Porcine circovirus ty
gb_vi:AF071879	- 231.00	498.94	8.6e-20	1758	AF071879 Porcine circovirus, c
gb_vi:AF012107	- 231.00	498.94	8.6e-20	1759	AF012107 Porcine circovirus, c
gb_vi:PCOMGEN	- 231.00	498.94	8.6e-20	1759	Y09921 Porcine circovirus comp
gb_vi:PCU49186	- 231.00	498.94	8.6e-20	1759	U49186 Porcine circovirus, com
gb_pr2:HUMARAAS	- 63.50	139.87	8.64	217	M30702 Human amphiregulin (AR)
gb_ba1:SSGM507	- 63.00	135.30	15.54	321	X79527 Streptococcus sp. group
gb_ba2:U67554	- 74.00	129.01	34.82	1029	U67554 Methanococcus jannasch
gb_pl2:AF005932	- 65.50	128.96	35.03	1299	AF005932 Clavispora opuntiae s
gb_in2:AF053904	- 62.50	126.60	47.42	758	AF053904 Ctenocephalides felis
gb_pl1:GHCA117A	- 58.00	126.48	48.15	243	Z27308 G. herbaceum (Al7A) DNA f
gb_com:MEBCASWR	- 64.50	126.13	50.34	1333	X54715 M. eugenii mRNA for beta
gb_vi:FLAPBIAL	- 66.00	124.55	61.65	2341	M35935 Influenza A/Korea/426/98
gb_sts:G24367	- 60.00	124.49	62.15	508	G24367 human STS WI-14184, sequ
gb_pl1:SCMTF2A	- 64.00	123.03	74.99	1669	X51665 S. cerevisiae mitochondr
gb_ba1:RP282310	- 60.00	122.78	77.41	617	Z83310 R. prowazekii genomic DNA
gb_ro:MMU38505	- 57.50	122.16	83.76	349	U38505 Mus musculus G protein b
gb_pl1:TMS8S	- 62.00	122.08	84.66	1114	X39991 T. baccata DNA for 5.8S
gb_pl1:MSCNAM1	- 64.00	121.66	89.40	1950	X14719 yeast NAM1 gene. 9/1993
gb_sts:G50499	- 58.00	121.47	91.53	429	G50499 SHGC-77977 Human Homo sa
em_pl1:AU066523	- 55.50	121.37	92.69	493	AU066523 Chlamydomonas sp. 18S
gb_pr2:HUMTCRBP	- 56.00	121.24	94.27	264	L34720 Homo sapiens T-cell rece
gb_ba2:UCU63355	- 55.00	121.22	94.58	205	U63355 Unidentified crenarchaeo
gb_pl1:ZMU17225	- 64.00	121.21	94.65	2051	U17225 Zea mays glucose-6 phos

gb\_pl1:SCYDL044C + 64.00 121.13 95.70 2071 ! 274092 S.cerevisiae chromos  
gb\_sts:G12367 + 57.00 120.52 103.39 370 ! G12367 Nile tilapia STS UNH2  
gb\_in1:AB013013 - 63.00 120.20 107.75 1781 ! AB013013 Doliolum nationali

seq\_name: gb\_vi:AF085695

seq\_documentation\_block:

LOCUS AF085695 1768 bp DNA circular VRL 30-SEP-1998  
DEFINITION Porcine circovirus strain 412, complete genome.

ACCESSION AF085695

NID 93668362

VERSION AF085695.1 GI:3668362

KEYWORDS porcine circovirus.

SOURCE Porcine circovirus

ORGANISM Viruses; ssDNA viruses; Circoviridae; Circovirus.

REFERENCE 1 (bases 1 to 1768)

AUTHORS Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.

TITLE Emergence of a new porcine circovirus

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1768)

AUTHORS Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.

TITLE Direct Submission

JOURNAL Submitted (21-AUG-1998) Veterinary Infectious Disease Organization,

University of Saskatchewan, 120 Veterinary Road, Saskatoon,

Saskatchewan S7N 5E3, Canada

FEATURES Location/Qualifiers

1..1768

/organism="porcine circovirus"

/strain="412"

/db\_xref="taxon:46221"

51..995

/codon\_start=1

/product="P35.8"

/protein\_id="AAC61860.1"

/db\_xref="PID:g3668363"

/db\_xref="GI:3668363"

/translation="MPSKNGRSQPQPKRWFTLNPSDEKRIKRLPISLDYFI

VGEENEGEPTPLHQLGFANFVKQTFNKVYLGARCHIEKAKGTDOQNKYCSKEGN

LLKECGAPSGQSGSDISTAVSTLSEGLTVAKQHPVTKVFRGAEILLKRVSGKM

LLKDGKTNVFIIVGPPCGSKWAANFANPETTYKPKNKWDGCHGKVVVDDFY

GWLPDLDLLRCRDYPLTVTKTGTVPLFARSILITSNOTPLEWYSSAVPAVEALYR

RITSLVFNKATQSTEGGQFVTLSPCPPEFYEN"

complement(357..671)

/codon\_start=1

/product="P12.1"

/protein\_id="AAC61865.1"

/db\_xref="PID:g3668368"

/db\_xref="GI:3668368"

/translation="MVTIPPLVFRWFPVCGFRVCKISSPFAFTTPRPHNEVYIGFPI

TLHFFPAHFQKFSQAPAEIFDKRYRVLICNGHNPALQQGTHSSRQVTPLSLRSSSTF

NK"

complement(386..565)

/codon\_start=1

/product="P6.6"

/protein\_id="AAC61866.1"

/db\_xref="PID:g3668369"

/db\_xref="GI:3668369"

/translation="MKCTLVFQSRFCIFPLTFKSSASPRKFLTNVTCGFATVTRIPL

SNKVLAVDRSLRCP"

553..732

/codon\_start=1

/product="P6.6"

/protein\_id="AAC61861.1"

/db\_xref="PID:g3668364"

/db\_xref="GI:3668364"

/translation="MVTSLMGHLGVKANGLLLIQTRKPTGNHLKTSGGMVTYVKKW

LLMTFMAGCRGIY"

833..970

/codon\_start=1

/product="P5.0"

/protein\_id="AAC61862.1"

/db\_xref="PID:g3668365"

FEATURES  
source

Location/Qualifiers  
1. 1768  
/organism="porcine circovirus"  
/strain="B9"  
/db\_xref="taxon:46221"  
51. 995  
/codon\_start=1  
/product="p35.8"  
/protein\_id="AAC61863.1"  
/db\_xref="PID:g3668366"  
/db\_xref="GI:3668366"  
/translation="MVFIHLGKVGKIFKSELYIRGYTDIVLVVFTVFERSAEA  
YVHISRLG"  
complement(1034. 1735)  
/codon\_start=1  
/product="p27.9"  
/protein\_id="AAC61864.1"  
/db\_xref="PID:g3668367"  
/db\_xref="GI:3668367"  
/translation="MTYPRRRYRRRRPRSHLQILRRRRLVHPHRYRWRRKNGI  
FNTLSRTFGTVKRTVTTPSWAVDMRFKIDDFVPGGNTNKISIPFEYIRIRKVK  
VEFWPCSPITQDGRGVGSTAVILDDNFVKATALTIDPYVNYSSRHTIPQPFYSHSRY  
FTPKPVLDSITIDYFQPNKRNQLRLQTSNVHDVHGLGTAFENSKYDQDYNIRVTMY  
VQFRENLKDPLEP"  
BASE COUNT 463 a 362 c 481 g 462 t  
ORIGIN

alignment\_scores:  
Quality: 311.00 Length: 59  
Ratio: 5.271 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-209-961-7 x AF085695/rev ..  
Align seg 1/1 to reverse of: AF085695 from: 1 to: 1768

1 MetLysCysThrLeuValPheGlnSerArgPheCysIlePheProLeuTh 17  
|||||  
565 ATGAAGTGATACATGGTTTCCCAATCACGCTTTTGCATTTCCCGCTCAC 516

17 rPheLysSerSerAlaSerProArgLysPheLeuThrAsnValThrGlyC 34  
|||||  
515 TTTCAAAGTTCAGCCAGCGCCGGGAAATTTTGACAAACGTTACAGGGT 466

34 yScysPheAlaThrValThrArgLysPheLeuThrAsnValThrGlyC 50  
|||||  
465 GCTGCTTTGCAACGGTCACCAAGTCCCGCTCCCAACAGGTACTCACA 416

51 AlavalAspArgSerLeuArgCysPro 59  
|||||  
415 GCAGTAGACAGGTCACTCCGTTGTCCT 389

seq\_name: gb\_vi:AF086834  
seq\_documentation\_block:  
LOCUS AF086834 1768 bp DNA circular VRL 29-SEP-1998  
DEFINITION Porcine circovirus strain B9, complete genome.  
ACCESSION AF086834  
NID 93661515  
VERSION AF086834.1 GI:3661518  
KEYWORDS  
SOURCE porcine circovirus.  
ORGANISM porcine circovirus.  
REFERENCE 1 (bases 1 to 1768)  
AUTHORS Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.  
TITLE Emergence of a new porcine circovirus  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1768)  
AUTHORS Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.  
TITLE Direct Submission  
JOURNAL Submitted (26-AUG-1998) VIDO, University of Saskatchewan, 120  
Veterinary Road, Saskatoon, Saskatchewan S7N 5E3, Canada

FEATURES  
source

Location/Qualifiers  
1. 1768  
/organism="porcine circovirus"  
/strain="B9"  
/db\_xref="taxon:46221"  
51. 995  
/codon\_start=1  
/product="p35.8"  
/protein\_id="AAC61737.1"  
/db\_xref="PID:g3661516"  
/db\_xref="GI:3661516"  
/translation="MPSKNGRSQGPQPHRWVFTLNNPSEDERKKIRELPISLDFYFI  
VVEEGNEGRTPHLOGFANFYKQTFNKVYKWLGAHIEKAGTDQDQNKYCSKEGN  
LLIECGAPRSQGRSGLSTAVTLLSEGLITVAKQHPVTYVKNFRGLAEKLVSGKM  
OKRDKNTVHFIVGPPGGCKSKAAANFANPETTYKPKKNKWDGYHGEKVVYDDFY  
GWLWDLLRLCDRYLPLVTKGTVPLARSILITSNQTPLEWYSSSTAVPAVEALYR  
RITSLVFWKNTQEQTGGQFVTLSPPCPEPFYEIN"  
complement(1034. 1735)  
/codon\_start=1  
/product="p27.9"  
/protein\_id="AAC61738.1"  
/db\_xref="PID:g3661517"  
/db\_xref="GI:3661517"  
/translation="MTYPRRRYRRRRPRSHLQILRRRRLVHPHRYRWRRKNGI  
FNTLSRTFGTVKRTVTTPSWAVDMRFKIDDFVPGGNTNKISIPFEYIRIRKVK  
VEFWPCSPITQDGRGVGSTAVILDDNFVKATALTIDPYVNYSSRHTIPQPFYSHSRY  
FTPKPVLDSITIDYFQPNKRNQLRLQTSNVHDVHGLGTAFENSKYDQDYNIRVTMY  
VQFRENLKDPLEP"  
BASE COUNT 463 a 357 c 482 g 466 t  
ORIGIN

alignment\_scores:  
Quality: 311.00 Length: 59  
Ratio: 5.271 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-209-961-7 x AF086834/rev ..  
Align seg 1/1 to reverse of: AF086834 from: 1 to: 1768

1 MetLysCysThrLeuValPheGlnSerArgPheCysIlePheProLeuTh 17  
|||||  
565 ATGAAGTGATACATGGTTTCCCAATCACGCTTTTGCATTTCCCGCTCAC 516

17 rPheLysSerSerAlaSerProArgLysPheLeuThrAsnValThrGlyC 34  
|||||  
515 TTTCAAAGTTCAGCCAGCGCCGGGAAATTTTGACAAACGTTACAGGGT 466

34 yScysPheAlaThrValThrArgLysPheLeuThrAsnValThrGlyC 50  
|||||  
465 GCTGCTTTGCAACGGTCACCAAGTCCCGCTCCCAACAGGTACTCACA 416

51 AlavalAspArgSerLeuArgCysPro 59  
|||||  
415 GCAGTAGACAGGTCACTCCGTTGTCCT 389

seq\_name: gb\_vi:AF086835  
seq\_documentation\_block:  
LOCUS AF086835 1768 bp DNA circular VRL 29-SEP-1998  
DEFINITION Porcine circovirus strain 9741, complete genome.  
ACCESSION AF086835  
NID 93661518  
VERSION AF086835.1 GI:3661518  
KEYWORDS  
SOURCE porcine circovirus.  
ORGANISM porcine circovirus.  
REFERENCE 1 (bases 1 to 1768)  
AUTHORS Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.  
TITLE Emergence of a new porcine circovirus

```
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1768)
AUTHORS Wang,L., Willson,P., Chow,B., Gibbons,E. and Babiuk,L.
TITLE Direct Submission
JOURNAL Submitted (26-AUG-1998) VIDO, University of Saskatchewan, 120
Veterinary Road, Saskatoon, Saskatchewan S7N 5E3, Canada
FEATURES
  source
    1..1768
      /organism="porcine circovirus"
      /strain="9741"
      /db_xref="taxon:46221"
      51..995
        /codon_start=1
        /product="p35, 8"
        /protein_id="AAC61739.1"
        /db_xref="PID:g3661519"
        /db_xref="GI:3661519"
      /translation="MPSKNGRSGPQPHKRWFTLNNPSENKRKKIRELPISLFDYFI
      VGEENEGRTPHLQGFANFVKOTENKVKWYLGARCHIEKAGTDQONKEVCSKEGN
      LIIECGAPRSQGSRLSTAVSILLESGLVTVAKQHPVTFVKNFRGLAELLKVSCKM
      QKRDWNTNHFIVGPGCGSKWAANFANPETTYWRPNKKNWDGTHGKVVVYIDDFY
      GWLPWDDLRLCDRYPLTVTKGTGTVPLARSILITSNOTPLEWYSSSTAVPAVEALYR
      RITSLVFNKATQSTEEGGQFVTLSPPCPEFFYEINY"
      complement(1034..1735)
      /codon_start=1
      /product="p27, 9"
      /protein_id="AAC61740.1"
      /db_xref="PID:g3661520"
      /db_xref="GI:3661520"
      /translation="MTYPRRRYRRRRHRSRHLGQILRRPWLVPVPHRYRWRKNGI
      ENTRLSRTFGYTVKRTVTPSWAVDMRFXIDDFVPGGNTKISIPFEYVIRKVK
      VEFWPCSPITQDGRGVSTAVILDNFTKATLYDYPYNYSSRHTIPQPFYSHSRY
      FTPKPLVDSTIDYFQPNKRNQLRLQTSNVDHVLGTAFENSKYDQDYNIRVTMY
      VQREFNLKDPPLP"
BASE COUNT 464 a 360 c 480 g 464 t
ORIGIN

alignment_scores:
  Quality: 311.00      Length: 59
  Ratio: 5.271        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-209-961-7 x AF086835/rev ..
  Align seg 1/1 to reverse of: AF086835 from: 1 to: 1768
    1 MetLysCysThrLeuValPheGlnSerArgPheCysIlePheProLeuTh 17
    565 ATGAAGTGATACATGGTTTCCAAATCACGCTTTTCATTTCCCGCTCAC 516
    17 rPheLysSerSerAlaSerProArgLysPheLeuThrAsnValThrGlyC 34
    515 TTTCAAAAGTTACAGCCCGCGGAAATTTTGACAAACGTTACAGGGT 466
    34 yScysPheAlaThrValThrArgIleProLeuSerAsnLysValLeuThr 50
    465 GCTGCTTTGCAACGGTCACCAAGATCCCGCTCCCAACAGGTACTCACA 416
    51 AlaValAspArgSerLeuArgCysPro 59
    415 GCAGTAGACAGGTCACTCGTGTGCTC 389
  seq_name: gb_vi:AF086836
  seq_documentation_block:
    LOCUS AF086836 1768 bp DNA circular VRL
    DEFINITION Porcine circovirus strain M226, complete genome.
    ACCESSION AF086836
    NID G3661521
    VERSION AF086836.1 GI:3661521
    KEYWORDS

SOURCE
  ORGANISM porcine circovirus
  Viruses; ssDNA viruses; Circoviridae; Circovirus.
REFERENCE 1 (bases 1 to 1768)
AUTHORS Wang,L., Willson,P., Chow,B., Gibbons,E. and Babiuk,L.
TITLE Emergence of a new porcine circovirus
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1768)
AUTHORS Wang,L., Willson,P., Chow,B., Gibbons,E. and Babiuk,L.
TITLE Direct Submission
JOURNAL Submitted (26-AUG-1998) VIDO, University of Saskatchewan, 120
Veterinary Road, Saskatoon, Saskatchewan S7N 5E3, Canada
FEATURES
  source
    1..1768
      /organism="porcine circovirus"
      /strain="M226"
      /db_xref="taxon:46221"
      51..995
        /codon_start=1
        /product="p35, 8"
        /protein_id="AAC61741.1"
        /db_xref="PID:g3661522"
        /db_xref="GI:3661522"
      /translation="MPSKNGRSGPQPHKRWFTLNNPSEDERKKIRELPISLFDYFI
      VGEENEGRTPHLQGFANFVKOTENKVKWYLGARCHIEKAGTDQONKEVCSKEGN
      LIIECGAPRSQGSRLSTAVSILLESGLVTVAKQHPVTFVKNFRGLAELLKVSCKM
      QKRDWNTNHFIVGPGCGSKWAANFANPETTYWRPNKKNWDGTHGKVVVYIDDFY
      GWLPWDDLRLCDRYPLTVTKGTGTVPLARSILITSNOTPLEWYSSSTAVPAVEALYR
      RITSLVFNKATQSTEEGGQFVTLSPPCPEFFYEINY"
      complement(1034..1735)
      /codon_start=1
      /product="p27, 9"
      /protein_id="AAC61742.1"
      /db_xref="PID:g3661523"
      /db_xref="GI:3661523"
      /translation="MTYPRRRYRRRRHRSRHLGQILRRPWLVPVPHRYRWRKNGI
      ENTRLSRTFGYTVKRTVTPSWAVDMRFXIDDFVPGGNTKISIPFEYVIRKVK
      VEFWPCSPITQDGRGVSTAVILDNFTKATLYDYPYNYSSRHTIPQPFYSHSRY
      FTPKPLVDSTIDYFQPNKRNQLRLQTSNVDHVLGTAFENSKYDQDYNIRVTMY
      VQREFNLKDPPLP"
BASE COUNT 463 a 363 c 480 g 462 t
ORIGIN

alignment_scores:
  Quality: 311.00      Length: 59
  Ratio: 5.271        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-209-961-7 x AF086836/rev ..
  Align seg 1/1 to reverse of: AF086836 from: 1 to: 1768
    1 MetLysCysThrLeuValPheGlnSerArgPheCysIlePheProLeuTh 17
    565 ATGAAGTGATACATGGTTTCCAAATCACGCTTTTCATTTCCCGCTCAC 516
    17 rPheLysSerSerAlaSerProArgLysPheLeuThrAsnValThrGlyC 34
    515 TTTCAAAAGTTACAGCCCGCGGAAATTTTGACAAACGTTACAGGGT 466
    34 yScysPheAlaThrValThrArgIleProLeuSerAsnLysValLeuThr 50
    465 GCTGCTTTGCAACGGTCACCAAGATCCCGCTCCCAACAGGTACTCACA 416
    51 AlaValAspArgSerLeuArgCysPro 59
    415 GCAGTAGACAGGTCACTCGTGTGCTC 389
  seq_name: gb_vi:AF118095
  seq_documentation_block:
```

LOCUS AF118095 597 bp DNA VRL 27-JUL-1999  
DEFINITION Porcine circovirus type 2 isolate IAF-614 unknown gene.  
ACCESSION AF118095  
NID 95596425  
VERSION AF118095.1 GI:5596425  
KEYWORDS porcine circovirus type 2.  
SOURCE Viruses: ssDNA viruses; Circoviridae; Circovirus.  
REFERENCE 1 (bases 1 to 597)  
AUTHORS Ouardani, M., Wilson, L., Jette, R., Montpetit, C. and Dea, S.  
TITLE Multiplex PCR for detection and typing of porcine circoviruses  
JOURNAL J. Clin. Microbiol. (1999) In press  
REFERENCE 2 (bases 1 to 597)  
AUTHORS Ouardani, M., Wilson, L., Jette, R., Montpetit, C. and Dea, S.  
TITLE Direct Submission  
JOURNAL Submitted (06-JAN-1999) CRMB, INRS-IAF, 531, Boulevard des  
Prairies, Laval, Que H7N 4Z3, Canada  
FEATURES  
source  
1..597  
/organism="porcine circovirus type 2"  
/isolate="IAF-614"  
/db\_xref="taxon:85708"  
/note="virus isolated from lung, lymph node, spleen and  
tonsil tissue from pigs affected by postweaning  
multisystemic wasting syndrome"  
CDS  
1..>597  
/note="ORF1"  
/codon\_start=1  
/product="unknown"  
/protein\_id="AAD4580.1"  
/db\_xref="PID:g5596426"  
/db\_xref="GI:5596426"  
/translation="SEDERKKIRLEPISLDFYFVIGESGNEGRTPHQQGFANFVKQ  
TFNKVKTGARGHIEKANGIDHONKEYSREGNLLIECAPRSOGQSDLSSTVSL  
LESGLVTAQHPVTFVRNFGRLAEKLVSGMKQKRDWKTNVHIVGPPGCGSKWA  
ANFADPTTYKPKPKWMDGYHGEVVIDDFYGLWLP"  
BASE COUNT 169 a 112 c 170 g 146 t  
ORIGIN  
alignment\_scores:  
Quality: 295.00 Length: 59  
Ratio: 5.175 Gaps: 0  
Percent Similarity: 96.610 Percent Identity: 94.915  
alignment\_block:  
US-09-209-961-7 x AF118095/rev ..  
Align seg 1/1 to reverse of: AF118095 from: 1 to: 597  
1 MetLysCysThrLeuValPheGlnSerArgPheCysIlePheProLeuTh 17  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
443 ATGACGTGTACATTGGTCTTCCCAATCAGCTTCTGCAATTTCCGCTCAC 394  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
17 rPheLysSerSerAlaSerProArgLysPheLeuThrAsnValThrGlyC 34  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
393 TTTCAAAAGTTCAGCCAGCCGCGGAAATTTCTGACAAACGTTACAGGTT 344  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
34 yScysPheAlaThrValThrArgIleProLeuSerAsnLysValLeuThr 50  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
343 GCTGCTCTGACAGGTCACAGACTCCCGCTCTCCACACAGGTTACTACA 294  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
51 AlaValAspArgSerLeuArgCysPro 59  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
293 GCAGTAGACAGGTCACCTCGTTGCTCT 267  
seq\_name: gb\_vi:AF055393  
seq\_documentation\_block:  
LOCUS AF055393 1767 bp DNA circular VRL 13-SEP-1998  
DEFINITION Porcine circovirus Type II from France, complete genome.  
ACCESSION AF055393

NID 93598820  
VERSION AF055393.1 GI:3598820  
KEYWORDS porcine circovirus.  
SOURCE Viruses: ssDNA viruses; Circoviridae; Circovirus.  
ORGANISM 1 (bases 1 to 1767)  
REFERENCE Meehan, B.M., McNeill, F., Todd, D., Kennedy, S., Jewhurst, V.A.,  
AUTHORS Ellis, J.A., Hassard, L.E., Clark, E.G., Haines, D.M. and Allan, G.M.  
TITLE Characterization of novel circovirus DNAs associated with wasting  
syndromes in pigs  
JOURNAL J. Gen. Virol. 79 (Pt 9), 2171-2179 (1998)  
MEDLINE 98418498  
REFERENCE 2 (bases 1 to 1767)  
AUTHORS Meehan, B.M.  
TITLE Direct Submission  
JOURNAL Submitted (26-MAR-1998) Brian M Meehan, Veterinary Science, The  
Queen's University of, Belfast, Stoney Road, Belfast, Northern  
Ireland, BT4 3SD, U.K.  
FEATURES  
Location/Qualifiers  
1..1767  
/organism="porcine circovirus"  
/specific\_host="Sus scrofa"  
/strain="Porcine circovirus Type II"  
/db\_xref="taxon:46221"  
/clone="Imp.1011 48121 EcORI No.17"  
/country="France"  
/complement(101..190)  
/codon\_start=1  
/product="ORF6"  
/protein\_id="AAC35329.1"  
/db\_xref="PID:g3598830"  
/db\_xref="GI:3598830"  
/translation="MASTPASPAPSDILSSSEPSQSRPPGRWT"  
103..210  
/codon\_start=1  
/product="ORF10"  
/protein\_id="AAC35327.1"  
/db\_xref="PID:g3598828"  
/db\_xref="GI:3598828"  
/translation="MSTAGVLTVVRLVYVVKRVRRLVKMPFFLLQR"  
261..320  
/codon\_start=1  
/product="ORF7"  
/protein\_id="AAC35328.1"  
/db\_xref="PID:g3598829"  
/db\_xref="GI:3598829"  
/translation="MAAGAVSSSPVTPPWIRHI"  
complement(311..439)  
/note="predicted 4.6 kda protein"  
/codon\_start=1  
/product="ORF9"  
/protein\_id="AAC35325.1"  
/db\_xref="PID:g3598826"  
/db\_xref="GI:3598826"  
/translation="MGLGSASSILLARHVAEVLPRCCRCRSALVILTAAHFFRFQI"  
398..1342  
/codon\_start=1  
/product="putative Rep protein (ORF1)"  
/protein\_id="AAC35320.1"  
/db\_xref="PID:g3598821"  
/db\_xref="GI:3598821"  
/translation="MPSKNGRSQPQPKRWFTLNPNSEDERKKIRDLISLDFYFI  
VEEGNEGRTPHQQGFANFVKQTENKVKLGARGCHIEKAKGTDQONKEYCSKEGN  
LIMECAPSQSQSDLSAVSLTLESGSLTVAEQHPVTFVRNFGRLAEKLVSGMK  
QKRDWKTNVHIVGPPGCGSKWAANFADPTTYKPKPKWMDGYHGEVVIDDFY  
GMLPWDDLRLICDRYPLTVETKGTVPFLARSLITSNQTPLEWYSSTAVPAVEALYR  
RITSLVFWKNAEQSTEEGGQFVTLSPPCPEFFFEYNY"  
674..679  
polyA\_signal  
complement(704..1018)  
/note="predicted 11.9 kda protein"  
/codon\_start=1  
/product="ORF3"

```
/protein_id="AAC35322.1"
/db_xref="PID:g3598823"
/db_xref="GI:3598823"
/translation="MTVLPVSRWFVCGFRVCKISSPFAFTPRPHNDVYLSLPI
TLHFFAHFQKFSQPAEISDKRYRLVLCNGHTPALQGGTHSSRQVTPLSRSRSTL
HQ"
complement(733..912)
/note="predicted 6.5 kDa protein"
/codon_start=1
/product="ORF4"
/protein_id="AAC35323.1"
/db_xref="PID:g3598824"
/db_xref="GI:3598824"
/translation="MTCTLVFQSRECFIFLTFKSSASPRKFLTNVTGCCSATVTRLPL
SNKVLTAVDRLRCP"
complement(1035..1100)
/note="predicted 2.3 kDa protein"
/codon_start=1
/product="ORF8"
/protein_id="AAC35324.1"
/db_xref="PID:g3598825"
/db_xref="GI:3598825"
/translation="MDIDHTVSDHPRAASHKSHQ"
1330..1335
complement(1336..1380)
/note="predicted 1.8 kDa protein"
/codon_start=1
/product="ORF11"
/protein_id="AAC35326.1"
/db_xref="PID:g3598827"
/db_xref="GI:3598827"
/translation="MNNKHVEYIKKTQ"
complement(1369..1374)
complement(join(1380..1767,1..314))
/note="predicted 27.8 kDa protein"
/codon_start=1
/product="ORF2"
/protein_id="AAC35321.1"
/db_xref="PID:g3598822"
/db_xref="GI:3598822"
/translation="MTYPRRYRRRRRSHLQILRRPWLPHRRHYNRKNGI
FNRLSTFGYTVKRTVTPSWADMMRNINDELPPGGSNPRSPFFYRIRKVK
VEWPCSPITQGRGVGSSAVILDNFTKATATYDPVYNSRHTITOPESVHSRY
FTPKPVLDSIDFQPNKRNLWLRLQTAGNDVHVLGTAFENSIYDQYENIRVTY
VQRENLKDPPLNP"
BASE COUNT 447 a 360 c 502 g 458 t
ORIGIN

alignment_scores:
  Quality: 295.00 Length: 59
  Ratio: 5.175 Gaps: 0
  Percent Similarity: 96.610 Percent Identity: 94.915

alignment_block:
US-09-209-961-7 x AF055393/rev ..

Align seg 1/1 to reverse of: AF055393 from: 1 to: 1767

1 MetLysCysThrLeuValPheGlnSerArgPheCysIlePheProLeuTh 17
|||||
912 ATGACGTGTACATTAGTCTTCAATACACGCTTCTGCATTTTCCCGCTCAC 863
|||||
17 rPheLysSerSerAlaSerProArgLysPheLeuThrAsnValThrGlyC 34
|||||
862 TTTCAAAGTTCAGCCAGCCCGCGGAATTTCTGACAAACGTTACAGGT 813
|||||
34 yScysPheAlaThrValThrArgIleProLeuSerAsnLysValLeuThr 50
|||||
812 GCTGCTCTCAACGGTCAACACACTCCCGCTCTCCAAACAGGTACTCACA 763
|||||
51 AlaValAspArgSerLeuArgCysPro 59
|||||
```

```
762 GCAGTAGACAGGTCACTCCGCTTCTCCC 736
seq_name: gb_vi:AF055394

seq_documentation_block:
LOCUS AF055394 1767 bp DNA circular VRL 13-SEP-1998
DEFINITION Porcine circovirus Type II from France, complete genome.
ACCESSION AF055394
NID g3598831
VERSION AF055394.1 GI:3598831
KEYWORDS
SOURCE
ORGANISM
porcine circovirus.
VIRUSES: ssDNA viruses; Circoviridae; Circovirus.
REFERENCE
1 (bases 1 to 1767)
AUTHORS
Meehan,B.M., McNeill,F., Todd,D., Kennedy,S., Jewhurst,V.A.,
Ellis,J.A., Hassard,L.E., Clark,E.G., Haines,D.M. and Allan,G.M.
TITLE
Characterization of novel circovirus DNAs associated with wasting
syndromes in pigs
J. Gen. Virol. 79 (Pt 9), 2171-2179 (1998)
JOURNAL
MEDLINE
98418498
AUTHORS
Meehan,B.M.
TITLE
Direct Submission
JOURNAL
Submitted (26-MAR-1998) Brian M Meehan, Veterinary Science, The
Queen's University of, Belfast, Stoney Road, Belfast, Northern
Ireland, BT4 3SD, U.K.
FEATURES
source
location/Qualifiers
1..1767
/organism="porcine circovirus"
/specific_host="Sus scrofa"
/strain="Porcine circovirus Type II"
/db_xref="taxon:46221"
/clone="Imp.1011 48285 EcoRI No.12"
/country="France"
/complement(101..190)
/codon_start=1
/product="ORF6"
/protein_id="AAC35339.1"
/db_xref="PID:g3598841"
/db_xref="GI:3598841"
/translation="MASSTPASPAPSDILSSPQSKRPDPGRWT"
103..210
/codon_start=1
/product="ORF10"
/protein_id="AAC35337.1"
/db_xref="PID:g3598839"
/db_xref="GI:3598839"
/translation="MSTAQEGVLTVVRLTVYPKVRERRVLMKMPFFLLQR"
261..320
/codon_start=1
/product="ORF7"
/protein_id="AAC35338.1"
/db_xref="PID:g3598840"
/db_xref="GI:3598840"
/translation="MAAGAVSSSPVTPPWIRHI"
complement(311..439)
/note="predicted 4.6 kDa protein"
/codon_start=1
/product="ORF9"
/protein_id="AAC35335.1"
/db_xref="PID:g3598837"
/db_xref="GI:3598837"
/translation="MGLGSASSILLAGHVAEVLPRCCRCRSALVILTAHFFRFQI"
398..1342
/codon_start=1
/product="putative Rep protein (ORF1)"
/protein_id="AAC35330.1"
/db_xref="PID:g3598832"
/db_xref="GI:3598832"
/translation="WPSKNGRSGPQPKRWVFTLNNPSEDERKKIRDLPSLFDYFI
VGESENGRTPHQLQGFANFVKQTFNKVKYLGARCHLEKAKGTDQONKEVCSKSGN
LMECGAPRSQGSDDLSTAVSLTLLSGSLTVAEQHPVTVFVNRFGLAELLKVSQGM
```



QKRDWNTNHHVIVGPPGCKSKWAANFADPTTYWKPNNKWDGTHGEVVDYDFY  
GWLWDDLLRLCDRPLTVETKGGTVPFLARSLITISNOTPLEWYSSTAVPAVEALYR  
RTSLVFNKNAEQSTEEGGQFVTLSPPCPEFFEYNY

polyA\_signal  
CDS

complement(704..1018)  
/note="predicted 11.9 kda protein"

/codon\_start=1  
/product="ORF3"

/protein\_id="AAC35332.1"

/db\_xref="PID:g3598834"

/db\_xref="GI:3598834"

/translation="MTYIPLVSRMFPVCGFRVCKISSFPAFTTPRPHNDVYISLPI  
TLHFPAHFQKESQPAEISDKRYVLLCNGHOTPALQCGTHSSRQVTPLSRSSTL  
H"

complement(733..912)

/note="predicted 6.5 kda protein"

/codon\_start=1

/product="ORF4"

/protein\_id="AAC35333.1"

/db\_xref="PID:g3598835"

/db\_xref="GI:3598835"

/translation="MTCTLVFQSRFCIFLTPFKSSAPRKFLTNVTGCGSNAVTRLPL  
SNKLVAVDRSLRCP"

complement(1035..1100)

/note="predicted 2.3 kda protein"

/codon\_start=1

/product="ORF8"

/protein\_id="AAC35334.1"

/db\_xref="PID:g3598836"

/db\_xref="GI:3598836"

/translation="MDIDTSTVSDHPRAASHKSHQ"

1330..1335

complement(1336..1380)

/note="predicted 1.8 kda protein"

/codon\_start=1

/product="ORF11"

/protein\_id="AAC35336.1"

/db\_xref="PID:g3598838"

/db\_xref="GI:3598838"

/translation="MNNKHVEVKKTKQ"

complement(1369..1374)

/note="predicted 27.8 kda protein"

/codon\_start=1

/product="ORF2"

/protein\_id="AAC35331.1"

/db\_xref="PID:g3598833"

/db\_xref="GI:3598833"

/translation="MTYPRRYRRRRPRSHLGOILRRPWLHVPHRYRWRKNGI  
FNRLSRFTGYTKFTVKTPTSWAVDMRFENDELPPGGSGNPRSPVPEYVIRKVK  
VFEPGSPITQDGRGVSSAVILDDNFVTKATALTDPVNTSSRRTIQTQPSYHSRY  
FTPKPVLDTIDYFQPNKRNQWLRLQTTGNVDHVLGTAFAFENSIYDQYINRVTMY  
VQREFNLKDPPLNP"

BASE COUNT 448 a 359 c 500 g 460 t

ORIGIN

alignment\_scores:

Quality: 295.00 Length: 59  
Ratio: 5.175 Gaps: 0  
Percent Similarity: 96.610 Percent Identity: 94.915

alignment\_block:

US-09-209-961-7 x AF055394/rev ..

Align seg 1/1 to reverse of: AF055394 from: 1 to: 1767

1 MetLysCysThrLeuValPheGlnSerArgPheCysIlePheProLeuThr 17

|||||

912 ATGACGGTGTACATTAGCTTCCATTCACGCTTCTGCAATTTCCCGCTCAC 863

|||||

17 rPheLysSerSerAlaSerProArgLysPheLeuThrAsnValThrGlyC 34

|||||

862 TTTCAAAAGTTTCAGCAGCCGCGGAAATTTCTGACAAAACGTTACAGGCT 813

34 yScysPheAlaThrValhrArgIleProLeuSerAsnLysValLeuThr 50

|||||

812 GCTGCTCTGCACAGGTCCACAGACTCCCGCTCTCCACAAGGTACTCACA 763

51 AlaValAspArgSerLeuArgCysPro 59

|||||

762 GCAGTAGACAGGTCACTCGGTGTGCC 736

seq\_name: gb\_v1:AF027217

seq\_documentation\_block:

LOCUS AF027217 1768 bp DNA circular VRL 14-MAY-1998

DEFINITION Porcine circovirus strain pmws PCV, complete genome.

ACCESSION AF027217

NID 92689645

VERSION AF027217.1 GI:2689645

KEYWORDS

SOURCE

ORGANISM

porcine circovirus

Viruses; ssDNA viruses; Circoviridae; Circovirus.

REFERENCE 1 (bases 1 to 1768)

AUTHORS Hamel,A.L., Lin,L.L. and Nayar,G.P.

TITLE Nucleotide sequence of porcine circovirus associated with

postweaning multisystemic wasting syndrome in pigs

J. Virol. 72 (6), 5262-5267 (1998)

MEDLINE 98241772

REFERENCE 2 (bases 1 to 1768)

AUTHORS Hamel,A.L., Lin,L.L. and Nayar,G.P.S.

TITLE Direct Submission

JOURNAL Submitted (26-SEP-1997) Virology Laboratory, Veterinary Services

Branch, Manitoba Agriculture, 545 University Crescent, Winnipeg,

Manitoba R3T 5S6, Canada

FEATURES

source

1..1768

/organism="porcine circovirus"

/strain="pmws PCV"

/db\_xref="taxon:46221"

/note="both strands of seven overlapping PCR fragments

were sequenced; virus isolated from lung, lymph node,

spleen and tonsil tissue from pigs affected by post

weaning multisystemic wasting syndrome"

51..995

/note="ORF1; similar to Rep protein encoded by

non-pathogenic PCV, GenBank Accession Number U49186;

predicted 35.8 kda protein"

/codon\_start=1

/product="putative Rep protein"

/protein\_id="AAC59462.1"

/db\_xref="PID:g2689646"

/db\_xref="GI:2689646"

/translation="MPSKNGRSQPQPKRWFTLNNPSEDEKKIRELPISLFDYFI

VGEENEGRTPHLGQFANFKVVKQTLNFKVILGARYIEKAKGTQDNKEYCSKEGN

LLTECGAPRSQGRSDLTAVSTLLESGSLVTAEOHPVTFVRNFRGLAELLKYSKGM

QKRDWNTNHHVIVGPPGCKSKWAANFADPTTYWKPNNKWDGTHGEVVDYDFY

GWLWDDLLRLCDRPLTVETKGGTVPFLARSLITISNOTPLEWYSSTAVPAVEALYR

RTSLVFNKNAEQSTEEGGQFVTLSPPCPEFFEYNY

117..125

/note="glycosylation site"

327..332

complement(357..671)

/note="ORF3; predicted 11.9 kda protein"

/codon\_start=1

/protein\_id="AAC59464.1"

/db\_xref="PID:g2689648"

/db\_xref="GI:2689648"

/translation="MTYIPLVSRMFPVCGFRVCKISSFPAFTTPRPHNDVYISLPI

TLHFPAHFQKESQPAEISDKRYVLLCNGHOTPALQCGTHSSRQVTPLSRSSTF

complement(386..565)

/note="ORF4; predicted 6.5 kda protein"

/codon\_start=1

```

/misc_feature
/db_xref="PID:g2689649"
/db_xref="GI:2689649"
/translation="WTCILVFSRRCIFPLTKSSASPRKFLNTVNGCCSATVTRLPL
SNKVLTAVDRLSRCP"
complement(470..478)
/note="glycosylation site"
complement(688..753)
/note="ORF8: predicted 2.3 kDa protein"
/codon_start=1
/protein_id="AAC59469.1"
/db_xref="PID:g2689653"
/db_xref="GI:2689653"
/translation="MDIDTVSDVHDPTAAASHKSHQ"
816..824
/note="glycosylation site"
906..914
/note="glycosylation site"
983..988
complement(989..1033)
/note="ORF11: predicted 1.8 kDa protein"
/codon_start=1
/protein_id="AAC59472.1"
/db_xref="PID:g2689656"
/db_xref="GI:2689656"
/translation="MKNKNHYEVIKKTQ"
1016..1177
/note="ORF5: predicted 6.2 kDa protein"
/codon_start=1
/protein_id="AAC59466.1"
/db_xref="PID:g2689650"
/db_xref="GI:2689650"
/translation="WVFPHLFGKWGVFKIKFSELYIHGYTDDIVLVLYVTVFERSAEA
YVHHISRGL"
complement(1022..1027)
complement(1034..1735)
/note="ORF2: predicted 27.8 kDa protein"
/codon_start=1
/protein_id="AAC59463.1"
/db_xref="PID:g2689647"
/db_xref="GI:2689647"
/translation="WTYPRRRYRRRRRSHLQQLRRRPLVHPRHRYRWRKNGI
FNTRLRFGYVTKATVTPSWAVDMRFNIDDFVPGGNTNKISIPPEYVIRKVK
VEFWCSPTQDGRGVSTAVILDNFVTKATALYDVPYVNSSRHTIPQPSYHSRY
FYPKPLVLSIDYFQPNNAKTQLWLRLQTSRNVDHVLGTAFENSIIYDQDINIRVTMI
VQRFNFKDPLKP"
complement(1301..1309)
/note="glycosylation site"
complement(1522..1611)
/note="ORF6: predicted 3.1 kDa protein"
/codon_start=1
/protein_id="AAC59467.1"
/db_xref="PID:g2689651"
/db_xref="GI:2689651"
/translation="MASSTPASPAPSDILSRLPQSRPPGRWT"
1524..1631
/note="ORF10: predicted 4.1 kDa protein"
/codon_start=1
/protein_id="AAC59471.1"
/db_xref="PID:g2689655"
/db_xref="GI:2689655"
/translation="MSTAQGLVTVVALVYPKVRRRVLKMPFFLLQR"
1682..1741
/note="ORF7: predicted 1.9 kDa protein"
/codon_start=1
/protein_id="AAC59468.1"
/db_xref="PID:g2689652"
/db_xref="GI:2689652"
/translation="MAAGAVSSAVTPPWIRHS"
complement(join(1732..1768,1..92))
/note="ORF9: predicted 4.6 kDa protein"
/codon_start=1
/protein_id="AAC59470.1"

```

```

/db_xref="PID:g2689654"
/db_xref="GI:2689654"
/translation="MWLGSSILLAGHVAEVLPRCCRCRSALVILTAHFFRFQL"
stem_loop
join(1746..1768,1..13)
/function="putative replication site"
rep_origin
join(1762..1768,1..2)
/note="AAGTATTAC" is similar to the nonanucleotide motif
in the non-pathogenic PCV, GenBank Accession Number
U49186"
BASE COUNT 452 a 360 c 495 g 461 t
ORIGIN

alignment_scores:
Quality: 295.00 Length: 59
Ratio: 5.175 Gaps: 0
Percent Similarity: 96.610 Percent Identity: 94.915

alignment_block:
US-09-209-961-7 x AF027217/rev ..

Align seg 1/1 to reverse of: AF027217 from: 1 to: 1768

1 MetLysCysThrLeuValPheGlnSerArgPheCysIlePheProLeuThr 17
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
565 ATGAGCGTGATACATTGGCTTCCAATCACGCTTCTGCAATTTCCCGCTCAC 516

17 rPheLysSerSerAlaSerProArgLysPheLeuThrAsnValThrGlyC 34
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
515 TTTCAAAAGTTACGACCGCCGCGGAAATTTCTGACAAACGTTACAGGCT 466

34 yScyPheAlaThrValThrArgIleProLeuSerAsnLysValLeuThr 50
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
465 GCTGCTGTCACAGGTCACGACCTCCGCTCTCCACAAGGTACTACACA 416

51 AlaValAspArgSerLeuArgCysPro 59
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
415 GCAGTAGACAGGTCACTCCGTGTGCTT 389

seq_name: gb_v1:AF055391

seq documentation_block:
LOCUS AF055391 1768 bp DNA circular VRL 13-SEP-1998
DEFINITION Porcine circovirus Type II from USA, complete genome.
ACCESSION AF055391
NID 93598796
VERSION AF055391.1 GI:3598796
KEYWORDS
SOURCE porcine circovirus.
ORGANISM porcine circovirus.
REFERENCE 1 (bases 1 to 1768)
Viruses; ssDNA viruses; Circoviridae; Circovirus.
AUTHORS Meehan,B.M., McNeilly,F., Todd,D., Kennedy,S., Jewhurst,V.A.,
Ellis,J.A., Hassard,L.E., Clark,E.G., Haines,D.M. and Allan,G.M.
TITLE Characterization of novel circovirus DNAs associated with wasting
syndromes in pigs
JOURNAL J. Gen. Virol. 79 (Pt 9), 2171-2179 (1998)
MEDLINE 98418498
REFERENCE 2 (bases 1 to 1768)
Meehan,B.M.
AUTHORS Direct Submission
TITLE Submitted (26-MAR-1998) Brian M Meehan, Veterinary Science, The
JOURNAL Queen's University of, Belfast, Stoney Road, Belfast, Northern
Ireland, BT4 3SD, U.K.
FEATURES
source
Location/Qualifiers
1..1768
/organism="porcine circovirus"
/specific_host="Sus scrofa"
/strain="Porcine circovirus Type II"
/db_xref="taxon:46221"
/clone="Imp. 999 EcoRI No.8"
/country="USA"
complement(101..190)
CDS

```

```

/codon_start=1
/product="ORF6"
/protein_id="AAC35308.1"
/db_xref="PID:g3598807"
/db_xref="GI:3598807"
/translation="MASSTPASPSDILSRLPQSERPPGRWT"
103..210
CDS

/codon_start=1
/product="ORF10"
/protein_id="AAC35306.1"
/db_xref="PID:g3598805"
/db_xref="GI:3598805"
/translation="MSTAQEGVLTVVALTVPKVRERRVLKMPFFLLQR"
261..320
CDS

/codon_start=1
/product="ORF7"
/protein_id="AAC35307.1"
/db_xref="PID:g3598806"
/db_xref="GI:3598806"
/translation="MAAGAVSSAVTTPPWIRHS"
complement(311..439)
CDS

/codon_start=1
/product="ORF9"
/protein_id="AAC35304.1"
/db_xref="PID:g3598803"
/db_xref="GI:3598803"
/translation="MWLGASASSILLAGHVAEVLPRCCRSALVILTAHFFRFQL"
398..1342
CDS

/codon_start=1
/product="putative Rep protein (ORF1)"
/protein_id="AAC35298.1"
/db_xref="PID:g3598797"
/db_xref="GI:3598797"
/translation="MPSKNGRSGPOPHKRWVFTLNNPSEDERKKIRELPISLFDYFI
VGEENEGRPHLOGFANFKQFNKYNWLGARCHEKAKGKTQQNKYCSKSGN
LLIEGAPRSQSDSLSTAVSLTESGLTVAEQHPVTFVFNFLGAEKLLKVSGRM
QKRMDLNVHIVGPPGCGSKWAANFADPETTKPPKRWKWDGYPHGEVVVIDDFY
GWLPMDDLRLCDRLPLVETKGGVFPFLARSILITSNQTPLEWYSSTAVPAVEALYR
RITSLVFWKATEQSTEGGQFVTLSPPCPEFYEIN"
674..679
complement(704..1018)
CDS

/codon_start=1
/product="predicted 11.9 kDa protein"
/protein_id="AAC35300.1"
/db_xref="PID:g3598799"
/db_xref="GI:3598799"
/translation="MVTIPLVSRNFPVCGFRVCKTSSPFAFTTPRPHNDVYIGLPI
TLLHPHAFKFSQPAEISDKRYVLLNGHQTALPQQGTHSSRQVTPLSLRSSSTF
NK"
complement(733..912)
CDS

/codon_start=1
/product="predicted 6.5 kDa protein"
/protein_id="AAC35301.1"
/db_xref="PID:g3598800"
/db_xref="GI:3598800"
/translation="MTCILVFQSRFCIFPLTFKSSASPRKFLTNVTCGSATVTRLPL
SNKVLTAVDRSLRCP"
complement(1035..1100)
CDS

/codon_start=1
/product="ORF8"
/protein_id="AAC35303.1"
/db_xref="PID:g3598802"
/db_xref="GI:3598802"
/translation="MDIDHTSVSDHPHTAASHKSHQ"
1330..1335
complement(1336..1380)
CDS

/codon_start=1
/product="predicted 1.8 kDa protein"
/protein_id="AAC35305.1"
/db_xref="PID:g3598804"
/db_xref="GI:3598804"
/translation="MNNKNHVEVIKKTQ"
1363..1524
CDS

/note="predicted 6.2 kDa protein"
/codon_start=1
/product="ORF5"
/protein_id="AAC35302.1"
/db_xref="PID:g3598801"
/db_xref="GI:3598801"
/translation="MVFIIHLGFKWGVFKIKFSELYIHGYTDIVLVVTVFERSAE"
1363..1524
CDS

/note="predicted 27.8 kDa protein"
/codon_start=1
/product="ORF2"
/protein_id="AAC35299.1"
/db_xref="PID:g3598798"
/db_xref="GI:3598798"
/translation="MTYPRRYRRRRHPRSHLQILRRRPLVHPHRYRWRRKNGI
ENTRLSFTFGYTVKATVTPTSWAVDMRNFIDDFVPGGNTNKISIPYEVYRIKKV
VEFWPCSPITQDRCVGSSTAVILDDNFVKATALYDPYVNTSSRTIQQPFYSHSRY
FTPKPVLSDTIDYFQPNKRNQLWLRLQTSRNVHDVGLGTAFENSYYDDYINIRVTMI
VQFREFNLKDPPLKP"
453 a 360 c 494 g 461 t
BASE COUNT
ORIGIN

alignment_scores:
Quality: 295.00 Length: 59
Ratio: 5.175 Gaps: 0
Percent Similarity: 96.610 Percent Identity: 94.915

alignment_block:
US-09-209-961-7 x AF055391/rev ..
Align seg 1/1 to reverse of: AF055391 from: 1 to: 1768
1 MetLysCysThrLeuValPheGlnSerArgPheCysIlePheProLeuTh 17
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
912 ATGACGTGTACATTTGGCTTCCCAATCACGCTTCGCAATTTCCCGCTCAC 863
17 rPheLysSerSerAlaSerProArgLysPheLeuThrAsnValThrGlyC 34
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
862 TTTCAAAGTTTCAGCCAGCCCGCGGAATTCGACAAACGTTACAGGT 813
34 yScysPheAlaThrValThrArgIleProLeuSerAsnLysValLeuThr 50
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
812 GCTGCTCTGCAACGGTCACAGACTCCGCTCTCCCAACAGGTACTCACA 763
51 AlaValAspArgSerLeuArgCysPro 59
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
762 GCAGTAGACAGGTCACTCCGTTGTCT 736
seq_name: gb_v1:AF055392
seq_documentation_block:
LOCUS AF055392 1768 bp DNA circular VRL 13-SEP-1998
DEFINITION Porcine circovirus Type II from Canada, complete genome.
ACCESSION AF055392
NID g3598808
VERSION AF055392.1 GI:3598808
KEYWORDS porcine circovirus.
SOURCE porcine circovirus
ORGANISM Viruses: ssDNA viruses; Circoviridae; Circovirus.
REFERENCE 1 (bases 1 to 1768)
AUTHORS Meehan,B.M., McNeill,F., Todd,D., Kennedy,S., Jewhurst,V.A.,
Ellis,J.A., Hassard,L.E., Clark,E.G., Haines,D.M. and Allan,G.M.
TITLE Characterization of novel circovirus DNAs associated with wasting
syndromes in pigs

```

J. Gen. Virol. 79 (Pt 9), 2171-2179 (1998)  
98418498  
2 (bases 1 to 1768)  
Meehan, B.M.  
Direct Submission  
Submitted (26-MAR-1998) Brian M Meehan, Veterinary Science, The  
Queen's University of, Belfast, Stoney Road, Belfast, Northern  
Ireland, BT4 3SD, U.K.  
Location/Qualifiers  
1. 1768  
/organism="porcine circovirus"  
/specific\_host="Sus scrofa"  
/strain="Porcine circovirus Type II"  
/db\_xref="taxon:46221"  
/clone="Imp.1010-Stoon EcoRI No.14"  
/country="Canada"  
complement(101..190)  
/codon\_start=1  
/product="ORF6"  
/protein\_id="AAC35319.1"  
/db\_xref="PID:g3598819"  
/db\_xref="GI:3598819"  
/translation="MASTPASPAPSDILSSVQSQRPGRWT"  
103..210  
/codon\_start=1  
/product="ORF10"  
/protein\_id="AAC35317.1"  
/db\_xref="PID:g3598817"  
/db\_xref="GI:3598817"  
/translation="MSTAQEGVTVVRLTVYPKVRERVLMKPFLLQR"  
261..320  
/codon\_start=1  
/product="ORF7"  
/protein\_id="AAC35318.1"  
/db\_xref="PID:g3598818"  
/db\_xref="GI:3598818"  
/translation="MAAGAVSSSANTPPWIRHS"  
complement(311..439)  
/note="predicted 4.6 kDa protein"  
/codon\_start=1  
/product="ORF9"  
/protein\_id="AAC35315.1"  
/db\_xref="PID:g3598815"  
/db\_xref="GI:3598815"  
/translation="MWLGASSILLAGHVAEVLPRCCRCRSALVILTAHFRFQL"  
398..1342  
/codon\_start=1  
/product="putative Rep protein (ORF1)"  
/protein\_id="AAC35309.1"  
/db\_xref="PID:g3598809"  
/db\_xref="GI:3598809"  
/translation="MPSKNGRSGPOPHKRWVTLNPNSEDERKKIRELPISLFDVFI  
VGESEGEETPHLQGFANFVKTKFNKYLGCARCHEKAKGTDOONKEYCSKGN  
LLIECGAPRQGSUDSTAVSLTESGLTVAEQHPVTFVFRNGLAELLKVSQK  
QKRWKTNHVIQVPPCGSKKANFADPETTYKPRKNKWDGHYGEVVDIDFY  
GWLPPDLLRLCTRYPTVETKGTVPFLARSILITNQPLEWISYTAIPAVALYR  
RITSLVFWKATEOSTEEGQFVTLSPCPPEFYNY"  
674..679  
complement(704..1018)  
/note="predicted 11.9 kDa protein"  
/codon\_start=1  
/product="ORF3"  
/protein\_id="AAC35311.1"  
/db\_xref="PID:g3598811"  
/db\_xref="GI:3598811"  
/translation="MVTIPPLVSMFVPCGFRVCKISSPFAFTTPRPHNDVYIGLPI  
TLHFPAHFKFSOPAEISDKRYRLVLCNGHQTALQOQGHSSRQVTPLSLRSSSTF  
NK"  
complement(733..912)  
/note="predicted 6.5 kDa protein"  
/codon\_start=1  
/product="ORF4"  
/protein\_id="AAC35312.1"

J. Gen. Virol. 79 (Pt 9), 2171-2179 (1998)  
98418498  
2 (bases 1 to 1768)  
Meehan, B.M.  
Direct Submission  
Submitted (26-MAR-1998) Brian M Meehan, Veterinary Science, The  
Queen's University of, Belfast, Stoney Road, Belfast, Northern  
Ireland, BT4 3SD, U.K.  
Location/Qualifiers  
1. 1768  
/organism="porcine circovirus"  
/specific\_host="Sus scrofa"  
/strain="Porcine circovirus Type II"  
/db\_xref="taxon:46221"  
/clone="Imp.1010-Stoon EcoRI No.14"  
/country="Canada"  
complement(101..190)  
/codon\_start=1  
/product="ORF6"  
/protein\_id="AAC35319.1"  
/db\_xref="PID:g3598819"  
/db\_xref="GI:3598819"  
/translation="MASTPASPAPSDILSSVQSQRPGRWT"  
103..210  
/codon\_start=1  
/product="ORF10"  
/protein\_id="AAC35317.1"  
/db\_xref="PID:g3598817"  
/db\_xref="GI:3598817"  
/translation="MSTAQEGVTVVRLTVYPKVRERVLMKPFLLQR"  
261..320  
/codon\_start=1  
/product="ORF7"  
/protein\_id="AAC35318.1"  
/db\_xref="PID:g3598818"  
/db\_xref="GI:3598818"  
/translation="MAAGAVSSSANTPPWIRHS"  
complement(311..439)  
/note="predicted 4.6 kDa protein"  
/codon\_start=1  
/product="ORF9"  
/protein\_id="AAC35315.1"  
/db\_xref="PID:g3598815"  
/db\_xref="GI:3598815"  
/translation="MWLGASSILLAGHVAEVLPRCCRCRSALVILTAHFRFQL"  
398..1342  
/codon\_start=1  
/product="putative Rep protein (ORF1)"  
/protein\_id="AAC35309.1"  
/db\_xref="PID:g3598809"  
/db\_xref="GI:3598809"  
/translation="MPSKNGRSGPOPHKRWVTLNPNSEDERKKIRELPISLFDVFI  
VGESEGEETPHLQGFANFVKTKFNKYLGCARCHEKAKGTDOONKEYCSKGN  
LLIECGAPRQGSUDSTAVSLTESGLTVAEQHPVTFVFRNGLAELLKVSQK  
QKRWKTNHVIQVPPCGSKKANFADPETTYKPRKNKWDGHYGEVVDIDFY  
GWLPPDLLRLCTRYPTVETKGTVPFLARSILITNQPLEWISYTAIPAVALYR  
RITSLVFWKATEOSTEEGQFVTLSPCPPEFYNY"  
674..679  
complement(704..1018)  
/note="predicted 11.9 kDa protein"  
/codon\_start=1  
/product="ORF3"  
/protein\_id="AAC35311.1"  
/db\_xref="PID:g3598811"  
/db\_xref="GI:3598811"  
/translation="MVTIPPLVSMFVPCGFRVCKISSPFAFTTPRPHNDVYIGLPI  
TLHFPAHFKFSOPAEISDKRYRLVLCNGHQTALQOQGHSSRQVTPLSLRSSSTF  
NK"  
complement(733..912)  
/note="predicted 6.5 kDa protein"  
/codon\_start=1  
/product="ORF4"  
/protein\_id="AAC35312.1"

J. Gen. Virol. 79 (Pt 9), 2171-2179 (1998)  
98418498  
2 (bases 1 to 1768)  
Meehan, B.M.  
Direct Submission  
Submitted (26-MAR-1998) Brian M Meehan, Veterinary Science, The  
Queen's University of, Belfast, Stoney Road, Belfast, Northern  
Ireland, BT4 3SD, U.K.  
Location/Qualifiers  
1. 1768  
/organism="porcine circovirus"  
/specific\_host="Sus scrofa"  
/strain="Porcine circovirus Type II"  
/db\_xref="taxon:46221"  
/clone="Imp.1010-Stoon EcoRI No.14"  
/country="Canada"  
complement(101..190)  
/codon\_start=1  
/product="ORF6"  
/protein\_id="AAC35319.1"  
/db\_xref="PID:g3598819"  
/db\_xref="GI:3598819"  
/translation="MASTPASPAPSDILSSVQSQRPGRWT"  
103..210  
/codon\_start=1  
/product="ORF10"  
/protein\_id="AAC35317.1"  
/db\_xref="PID:g3598817"  
/db\_xref="GI:3598817"  
/translation="MSTAQEGVTVVRLTVYPKVRERVLMKPFLLQR"  
261..320  
/codon\_start=1  
/product="ORF7"  
/protein\_id="AAC35318.1"  
/db\_xref="PID:g3598818"  
/db\_xref="GI:3598818"  
/translation="MAAGAVSSSANTPPWIRHS"  
complement(311..439)  
/note="predicted 4.6 kDa protein"  
/codon\_start=1  
/product="ORF9"  
/protein\_id="AAC35315.1"  
/db\_xref="PID:g3598815"  
/db\_xref="GI:3598815"  
/translation="MWLGASSILLAGHVAEVLPRCCRCRSALVILTAHFRFQL"  
398..1342  
/codon\_start=1  
/product="putative Rep protein (ORF1)"  
/protein\_id="AAC35309.1"  
/db\_xref="PID:g3598809"  
/db\_xref="GI:3598809"  
/translation="MPSKNGRSGPOPHKRWVTLNPNSEDERKKIRELPISLFDVFI  
VGESEGEETPHLQGFANFVKTKFNKYLGCARCHEKAKGTDOONKEYCSKGN  
LLIECGAPRQGSUDSTAVSLTESGLTVAEQHPVTFVFRNGLAELLKVSQK  
QKRWKTNHVIQVPPCGSKKANFADPETTYKPRKNKWDGHYGEVVDIDFY  
GWLPPDLLRLCTRYPTVETKGTVPFLARSILITNQPLEWISYTAIPAVALYR  
RITSLVFWKATEOSTEEGQFVTLSPCPPEFYNY"  
674..679  
complement(704..1018)  
/note="predicted 11.9 kDa protein"  
/codon\_start=1  
/product="ORF3"  
/protein\_id="AAC35311.1"  
/db\_xref="PID:g3598811"  
/db\_xref="GI:3598811"  
/translation="MVTIPPLVSMFVPCGFRVCKISSPFAFTTPRPHNDVYIGLPI  
TLHFPAHFKFSOPAEISDKRYRLVLCNGHQTALQOQGHSSRQVTPLSLRSSSTF  
NK"  
complement(733..912)  
/note="predicted 6.5 kDa protein"  
/codon\_start=1  
/product="ORF4"  
/protein\_id="AAC35312.1"

J. Gen. Virol. 79 (Pt 9), 2171-2179 (1998)  
98418498  
2 (bases 1 to 1768)  
Meehan, B.M.  
Direct Submission  
Submitted (26-MAR-1998) Brian M Meehan, Veterinary Science, The  
Queen's University of, Belfast, Stoney Road, Belfast, Northern  
Ireland, BT4 3SD, U.K.  
Location/Qualifiers  
1. 1768  
/organism="porcine circovirus"  
/specific\_host="Sus scrofa"  
/strain="Porcine circovirus Type II"  
/db\_xref="taxon:46221"  
/clone="Imp.1010-Stoon EcoRI No.14"  
/country="Canada"  
complement(101..190)  
/codon\_start=1  
/product="ORF6"  
/protein\_id="AAC35319.1"  
/db\_xref="PID:g3598819"  
/db\_xref="GI:3598819"  
/translation="MASTPASPAPSDILSSVQSQRPGRWT"  
103..210  
/codon\_start=1  
/product="ORF10"  
/protein\_id="AAC35317.1"  
/db\_xref="PID:g3598817"  
/db\_xref="GI:3598817"  
/translation="MSTAQEGVTVVRLTVYPKVRERVLMKPFLLQR"  
261..320  
/codon\_start=1  
/product="ORF7"  
/protein\_id="AAC35318.1"  
/db\_xref="PID:g3598818"  
/db\_xref="GI:3598818"  
/translation="MAAGAVSSSANTPPWIRHS"  
complement(311..439)  
/note="predicted 4.6 kDa protein"  
/codon\_start=1  
/product="ORF9"  
/protein\_id="AAC35315.1"  
/db\_xref="PID:g3598815"  
/db\_xref="GI:3598815"  
/translation="MWLGASSILLAGHVAEVLPRCCRCRSALVILTAHFRFQL"  
398..1342  
/codon\_start=1  
/product="putative Rep protein (ORF1)"  
/protein\_id="AAC35309.1"  
/db\_xref="PID:g3598809"  
/db\_xref="GI:3598809"  
/translation

seq_name: gb_vi:AF109397	seq_documentation_block:	1768 bp	DNA circular	VRL	02-FEB-1999
LOCUS	AF109397	1768 bp	DNA circular	VRL	02-FEB-1999
DEFINITION	Bovine circovirus, complete genome.				
ACCESSION	AF109397				
NID	94210313				
VERSION	AF109397.1	GI:4210313			
KEYWORDS					
SOURCE	bovine circovirus.				
ORGANISM	Viruses; ssDNA viruses; Circoviridae; Circovirus.				
REFERENCE	1 (bases 1 to 1768)				
AUTHORS	Hamel,A.L. and Nayar,G.P.S.				
TITLE	Submitted (27-NOV-1998) Virology Laboratory, Veterinary Diagnostic				
JOURNAL	Laboratory, Manitoba Agriculture, 545 University Crescent,				
REFERENCE	Winnipeg, Manitoba R3T 5S6, Canada				
AUTHORS	3 (bases 1 to 1768)				
TITLE	Hamel,A.L. and Nayar,G.P.S.				
JOURNAL	Submitted (02-FEB-1999) Virology Laboratory, Veterinary Diagnostic				
REFERENCE	Laboratory, Manitoba Agriculture, 545 University Crescent,				
AUTHORS	Winnipeg, Manitoba R3T 5S6, Canada				
REMARK	Sequence update by submitter				
COMMENT	On Feb 2, 1999 this sequence version replaced gi:4028609.				
FEATURES	Location/Qualifiers				
source	1..1768				
	/organism="bovine circovirus"				
	/virus				
	/db_xref="taxon:85542"				
	/note="BCV				
	sequence was obtained from several overlapping PCRs using				
	DNA extracted from various tissues from cattle; similar to				
	Porcine circovirus sequence presented in GenBank Accession				
	Number AF027217"				
repeat_region	13..35				
	/rpt_type=tandem				
	/rpt_unit=13..18				
CDS	51..995				
	/note="ORF-1; similar to Porcine circovirus ORF1 encoded				
	by the sequence presented in GenBank Accession Number				
	AF027217."				
	/codon_start=1				
	/product="putative Rep and coat protein"				
	/protein_id="AAD11928.1"				
	/db_xref="PID:g4210314"				
	/db_xref="GI:4210314"				
	/translation="MPSKKNGRSGPQPKRWFTLNNPSEDERKKIHELPISEDFEI				
	VLEEGEGRPHLQGFANFVKQTFNFKWILGARCHIEKAKGTQQNKEYCSKEGN				
	LLIECGPRSGQSDLSSTAVSTLLESGSLVTVAEQHPVTFVFNRLAEKLVSGKM				
	QKRDKWTNVHIVGPPCGSKAANFADPETYTWKPKRWKWDGHHGEVVIDFY				
	GWLWDLLRLCDRLPTVETKGTVPFLARSILITSPLEWISSSTAVPAVEALYR				
	RITSILVKNKATESTEEGGFVTLSPCPPEFYENY"				
	327..332				
	complement(357..671)				
	/note="similar to Porcine circovirus ORF-3 encoded by the				
	sequence presented in GenBank Accession Number AF027217."				
	/codon_start=1				
	/product="ORF-3"				
	/protein_id="AAD11930.1"				
	/db_xref="PID:g4210315"				
	/db_xref="GI:4210315"				
	/translation="MVTIPPLVSRWFVCGFRWCKISSPFAFTPRWPHNDVYIGLPI				
	TLLHPPAFQFQSPAISDKRYRLVLCNGHQPALQQGTHSSRQVTLPSRSSTFF				
	NK"				
	complement(386..565)				
	/note="similar to Porcine circovirus ORF-4 encoded by the				
	sequence presented in GenBank Accession Number AF027217."				
	/codon_start=1				
	/product="ORF-4"				
	/protein_id="AAD11931.1"				
	/db_xref="PID:g4210316"				
	/db_xref="GI:4210316"				
	/translation="MTCTLVFQSRFCIFPLTFKSSASPRKFLNVTGCCSATVTRPLP				
	SNKVLTAVDNLSRCP"				
	553..732				
	/codon_start=1				
	/product="ORF-12"				
	/protein_id="AAD11927.1"				
	/db_xref="PID:g4028610"				
	/db_xref="GI:4028610"				
	/translation="MYTSLMGLGVKANGLLILQTRKPHGTGNHLETSGGNVTVKRWK				
	LLMTFMAGCRGMiy"				
	complement(688..753)				
	/codon_start=1				
	/product="ORF-8"				
	/protein_id="AAD11935.1"				
	/db_xref="PID:g4028618"				
	/db_xref="GI:4028618"				
	/translation="MDIDHTVSDVDPHTAASHKSHQ"				
	983..988				
	complement(989..1033)				
	/codon_start=1				
	/product="ORF-11"				
	/protein_id="AAD11937.1"				
	/db_xref="PID:g4028620"				
	/db_xref="GI:4028620"				

PF EY R I R K V K

```

VEFWPCSPITQDGRGVSSAILDDNFVIRKATQYDPYVNYSSRHTIPOPFYSHSRY
FTFPVLDSITIDYFQPNKRNQLWRLQTSRNVHVLGTAFENSKYDQDYNIRVTMY
VQREFNLKDPPLKP"
1524..1631
/codon_start=1
/product="ORF-10"
/protein_id="AAD03079.1"
/db_xref="PID:g4106902"
/db_xref="GI:4106905"
/translation="MSTAQEGVLTVVTPKVRERRALKMPFLFOR"
complement(1528..1611)
/codon_start=1
/product="ORF-6"
/protein_id="AAD03076.1"
/db_xref="PID:g4106902"
/db_xref="GI:4106902"
/translation="MASSTPASPPSILSLRLOSARPPGR"
1682..1741
/codon_start=1
/product="ORF-7"
/protein_id="AAD03077.1"
/db_xref="PID:g4106903"
/db_xref="GI:4106903"
/translation="MAAGAVSSSAETPPWIRHS"
join(1750..1768,1..13)
/note="putative; similar to the nonanucleotide motif of
Porcine circovirus presented in GenBank Accession Numbers
AF027217 and U49186"
BASE COUNT 448 a 362 c 495 g 463 t
ORIGIN

alignment_scores:
  Quality: 295.00      Length: 59
  Ratio: 5.175        Gaps: 0
  Percent Similarity: 96.610      Percent Identity: 94.915

alignment_block:
US-09-209-961-7 x AF109399/rev ..
Align seg 1/1 to reverse of: AF109399 from: 1 to: 1768
1 MetLysCysThrLeuValPheGlnSerArgPheCysIlePheProLeuThr 17
|||||
565 ATGACGTGTACATTGGCTCTCCAAATTCACGCTTCTGCAATTTCCCGCTCAC 516

17 rPheLysSerAlaSerProArgLysPheLeuThrAsnValThrGlyc 34
|||||
515 TTTCAAAGTTCAGCCAGCCGCGGAAATTTCTGACAAACGTTACAGGT 466

34 yscCysPheAlaThrValThrArgIleProLeuSerAsnLysValLeuThr 50
|||||
465 GCTGCTCTCCAAAGTTCAGCCAGCTCCCGCTCTCCAAACAGGTACTACACA 416

51 AlaValAspArgSerLeuArgCysPro 59
|||||
415 GCAGTAGACAGGTACTCCGTTGTCTCT 389

seq_name: gb_v1:AF147751

seq_documentation_block:
LOCUS AF147751 1768 bp DNA VRL 08-JUN-1999
DEFINITION Porcine circovirus type 2 unknown genes.
ACCESSION AF147751
NID 95007010
VERSION AF147751.1 GI:5007010
KEYWORDS porcine circovirus type 2.
SOURCE porcine circovirus type 2.
ORGANISM porcine circovirus type 2
REFERENCE 1 (bases 1 to 1768)
AUTHORS Pogranichnyy, R., Yoon, K.-J., Harms, P., Swenson, S., Zimmerman, J. and

```

```

Sorden, S.
Characterization of clinical and immune responses in young swine to
experimental porcine circovirus type II infection
Unpublished
2 (bases 1 to 1768)
Yoon, K.-J. and Pogranichnyy, R.M.
Direct Submission
Submitted (30-APR-1999) Veterinary Diagnostic Laboratory, Iowa
State University, College of Veterinary Medicine, Ames, Iowa 50011,
USA
FEATURES
  source
    1..1768
      /organism="porcine circovirus type 2"
      /isolate="ISUVDL 98-15237"
      /specific_host="Sus scrofa"
      /db_xref="taxon:83708"
      /note="Isolated from pigs with postweaning multisystemic
      wasting syndrome"
      51..995
        /note="ORF1: 35-36kD"
        /codon_start=1
        /product="unknown"
        /protein_id="AAD37776.1"
        /db_xref="PID:g5007011"
        /db_xref="GI:5007011"
        /translation="MPSKNGRSQPPKRWVFTLNNPSEDERKKIRELPISLDFYFI
        VGESENGRTPHLQGFANFYKQTFNKVYLGARCHKIERAKGTDOONKEYCSKEGN
        LLIEGAPRSQGRSDLSSTAVSLLESGLTVAEQHPVTFVRNFRGLAELLKVGKMK
        QKRWKTNHVIVGPGGCKSKWAANFADPETTYWKPPRNKWDGTHGEVAVVDDFY
        GWLPDLDRLCDRYPVTETKGTGTPFLARSILITSNQTPLEWYSSTAAPVAEALYR
        RITSLVFNKNAEQSTERGQFVTLSPCPPEFFYEINY"
        complement(1034..1735)
        /note="ORF2: 27-28kD"
        /codon_start=1
        /product="unknown"
        /protein_id="AAD37777.1"
        /db_xref="PID:g5007012"
        /db_xref="GI:5007012"
        /translation="MTYPRRYRRHRPRSHLGOILRRPWLVPVHPRHYRWRKNGI
        ENTRLSRFGTGTVKATVTFSTSWAVDMRNFIDDFVPGGKTNKISIFEYTRIRKVK
        VEFWPCSPITQDGRGVSSAILDDNFVIRKATQYDPYVNYSSRHTIPOPFYSHSRY
        FTFPVLDSITIDYFQPNKRNQLWRLQTSRNVHVLGTAFENSKYDQDYNIRVTMY
        VQREFNLKDPPLKP"
BASE COUNT 450 a 362 c 494 g 462 t
ORIGIN

alignment_scores:
  Quality: 295.00      Length: 59
  Ratio: 5.175        Gaps: 0
  Percent Similarity: 96.610      Percent Identity: 94.915

alignment_block:
US-09-209-961-7 x AF147751/rev ..
Align seg 1/1 to reverse of: AF147751 from: 1 to: 1768
1 MetLysCysThrLeuValPheGlnSerArgPheCysIlePheProLeuThr 17
|||||
565 ATGACGTGTACATTGGCTCTCCAAATTCACGCTTCTGCAATTTCCCGCTCAC 516

17 rPheLysSerAlaSerProArgLysPheLeuThrAsnValThrGlyc 34
|||||
515 TTTCAAAGTTCAGCCAGCCGCGGAAATTTCTGACAAACGTTACAGGT 466

34 yscCysPheAlaThrValThrArgIleProLeuSerAsnLysValLeuThr 50
|||||
465 GCTGCTCTCCAAAGTTCAGCCAGCTCCCGCTCTCCAAACAGGTACTACACA 416

51 AlaValAspArgSerLeuArgCysPro 59
|||||
415 GCAGTAGACAGGTACTCCGTTGTCTCT 389

```

seq\_name: gb\_vi:AF154679

seq\_documentation\_block: 1768 bp DNA circular VRL 13-JUN-1999  
 LOCUS AF154679 Porcine circovirus complete genome.  
 DEFINITION Porcine circovirus, complete genome.  
 ACCESSION AF154679  
 NID 95052004  
 VERSION AF154679.1 GI:5052004  
 KEYWORDS porcine circovirus.  
 SOURCE  
 ORGANISM  
 Viruses; ssDNA viruses; Circoviridae; Circovirus.  
 REFERENCE 1 (bases 1 to 1768)  
 AUTHORS Kuo,T.Y., Chiou,Y.C. and Lai,S.S.  
 TITLE Complete nucleotide sequences analysis of porcine circovirus outbreak in Taiwan  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 1768)  
 AUTHORS Kuo,T.Y., Chiou,Y.C. and Lai,S.S.  
 TITLE Direct Submission  
 JOURNAL Submitted (27-MAY-1999) Veterinary Medicine, National Taiwan University, 142, Chouan Road, Taipei 106, Taiwan

FEATURES  
 source  
 1..1768  
 /organism="porcine circovirus"  
 /isolate="MLT98"  
 /db\_xref="taxon:46221"  
 /country="Taiwan"  
 51..995  
 /notes="putative Rep protein"  
 /codon\_start=1  
 /product="ORF1"  
 /protein\_id="AAD38398.1"  
 /db\_xref="PID:g5052005"  
 /db\_xref="GI:5052005"  
 /translation="MPSKNGRSGPQPKRWFTLNPSDEKRRKIRELPISLFDCTI  
 VGESENEGRTPHLQGFANFKVKTENKVKYFGARCHIEKAGTDOONKEYCSKEG  
 LLIECAPRSQGRSDLSSTAVSLLESGLVTVAEQHPVTVFRNFRGLAELLKVSCKM  
 QKDKWTNHHVIVGPPCGSKWAANFADPETTYWPKRNKWDGHHGEVVDFFY  
 GWLPDLDLRCLDPLVTETKGTVPFLARSILITSNQTPLEWYSSTAVPAVEALYR  
 RITSLVFWKNATQSTEGGQFVTLSPCPPEFFIYNY"

CDS

BASE COUNT 451 a 367 c 495 g 455 t  
 ORIGIN

alignment\_scores:  
 Quality: 295.00 Length: 59  
 Ratio: 5.175 Gaps: 0  
 Percent Similarity: 96.610 Percent Identity: 94.915

alignment\_block:  
 US-09-209-961-7 x AF154679/rev ..

Align seg 1/1 to reverse of: AF154679 from: 1 to: 1768

1 MetLysCysThrLeuValPheGlnSerArgPheCysIlePheProLeuTh 17  
 ||| ||||||||||||||||||||||||||||||||||||||||||||  
 565 ATGACGTGTACATTGGTCTTCCAATCAGCTTCTGCAATTTCCCGCTCAC 516  
 17 rPhelysSerSerAlaSerProArgLysPheLeuThrAsnValThrGlyC 34  
 ||| ||||||||||||||||||||||||||||||||||||||||||||  
 515 TTTCAAAAGTTTCAGCAGCCCGCGGAAATTTCTGACAAACGTTACAGGT 466  
 34 yScysPheAlaThrValThrArgIleProLeuSerAsnLysValLeuThr 50  
 ||| ||||||||||||||||||||||||||||||||||||||||||||  
 465 GCTGCTCTGCACAGGTACACGACTCCCGCTCTCCACAGGTACTCACA 416  
 51 AlaValAspArgSerLeuArgCysPro 59  
 ||| ||||||||||||||||||||||||||||||||||||||||||||  
 415 GCAGTAGACAGGTCACTGCTGTGCTCT 389

seq\_name: gb\_vi:AF166528

seq\_documentation\_block: 1768 bp mRNA VRL 19-AUG-1999  
 LOCUS AF166528 Porcine circovirus complete genome.  
 DEFINITION Porcine circovirus complete genome.  
 ACCESSION AF166528  
 NID 95739338  
 VERSION AF166528.1 GI:5739338  
 KEYWORDS porcine circovirus.  
 SOURCE  
 ORGANISM  
 Viruses; ssDNA viruses; Circoviridae; Circovirus.  
 REFERENCE 1 (bases 1 to 1768)  
 AUTHORS Yang,K.H., Lee,Y.F., Chao,D.S., Shieh,Y.C. and Lai,S.S.  
 TITLE Complete nucleotide sequences of porcine circovirus Tainan strand outbreak in Taiwan  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 1768)  
 AUTHORS Yang,K.H., Lee,Y.F., Chao,D.S., Shieh,Y.C. and Lai,S.S.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-JUL-1999) Veterinary Medicine, National Chia-Yi Institute of Technology, 300 Shei Fu Road, Lu Liao Li, Chia-Yi City 600, Taiwan

FEATURES  
 source  
 1..1768  
 /organism="porcine circovirus"  
 /virion  
 /strain="Tainan"  
 /db\_xref="taxon:46221"  
 /country="Taiwan"  
 51..995  
 /codon\_start=1  
 /product="putative Rep protein"  
 /protein\_id="AAD50432.1"  
 /db\_xref="PID:g5739339"  
 /db\_xref="GI:5739339"  
 /translation="MPSKNGRSGPQPKRWFTLNPSDEKRRKIRELPISLFDYFI  
 VGESENEGRTPHLQGFANFKVKTENKVKYFGARCHIEKAGTDOONKEYCSKEG  
 LLIECAPRSQGRSDLSSTAVSLLESGLVTVAEQHPVTVFRNFRGLAELLKVSCKM  
 QKDKWTNHHVIVGPPCGSKWAANFADPETTYWPKRNKWDGHHGEVVDFFY  
 GWLPDLDLRCLDPLVTETKGTVPFLARSILITSNQTPLEWYSSTAVPAVEALYR  
 RITSLVFWKNATQSTEGGQFVTLSPCPPEFFIYNY"

CDS

BASE COUNT 453 a 367 c 492 g 456 t  
 ORIGIN

alignment\_scores:  
 Quality: 295.00 Length: 59  
 Ratio: 5.175 Gaps: 0  
 Percent Similarity: 96.610 Percent Identity: 94.915

alignment\_block:  
 US-09-209-961-7 x AF166528/rev ..

Align seg 1/1 to reverse of: AF166528 from: 1 to: 1768

1 MetLysCysThrLeuValPheGlnSerArgPheCysIlePheProLeuTh 17  
 ||| ||||||||||||||||||||||||||||||||||||||||||||  
 565 ATGACGTGTACATTGGTCTTCCAATCAGCTTCTGCAATTTCCCGCTCAC 516  
 17 rPhelysSerSerAlaSerProArgLysPheLeuThrAsnValThrGlyC 34  
 ||| ||||||||||||||||||||||||||||||||||||||||||||  
 515 TTTCAAAAGTTTCAGCAGCCCGCGGAAATTTCTGACAAACGTTACAGGT 466  
 34 yScysPheAlaThrValThrArgIleProLeuSerAsnLysValLeuThr 50  
 ||| ||||||||||||||||||||||||||||||||||||||||||||  
 465 GCTGCTCTGCACAGGTACACGACTCCCGCTCTCCACAGGTACTCACA 416  
 51 AlaValAspArgSerLeuArgCysPro 59  
 ||| ||||||||||||||||||||||||||||||||||||||||||||  
 415 GCAGTAGACAGGTCACTGCTGTGCTCT 389



---

OM of: US-09-209-961-7 to: N\_Geneseq\_36:\* out\_format : pfs  
Date: Dec 28, 1999 1:50 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:  
-MODEL=framet\_p2n.model -DEV=xlp  
-Q/cgn2.1/USPIO.spool/US0209961/runat\_22121999\_101043\_17942/app\_query.fasta.1  
-DB=N\_Geneseq\_36 -OFMT=fastap -SUFFIX=ring -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-GAPOP=6.000 -CGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-DELOP=6.000 -FCAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=escore  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09209961  
-NCPU=6 -ICPU=3 -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-209-961-7  
Query length: 59  
Database: N\_Geneseq\_36:\*  
Database sequences: 311585  
Database length: 125096042  
Search time (sec): 140.030000

seq_list:	Sequence	Strid	Orig	ZScore	EScore	Len	Documentation
N_Geneseq_36:X35210	-	295.00	705.57	1.1e-31	1767	1767	Nucleotide sequence of PCV iso
N_Geneseq_36:X35011	-	295.00	705.57	1.1e-31	1767	1767	Nucleotide sequence of PCV iso
N_Geneseq_36:X35012	-	295.00	705.56	1.1e-31	1768	1768	Genomic DNA sequence of PCV st
N_Geneseq_36:X35213	-	295.00	705.56	1.1e-31	1768	1768	Nucleotide sequence of PCV iso
N_Geneseq_36:X35212	-	295.00	705.56	1.1e-31	1768	1768	Nucleotide sequence of PCV iso
N_Geneseq_36:X35214	-	295.00	705.56	1.1e-31	1768	1768	Nucleotide sequence of PCV iso
N_Geneseq_36:X35013	+	231.00	546.01	8.2e-23	1759	1759	Nucleotide sequence which has
N_Geneseq_36:X35012	-	69.50	150.11	0.9245	925	925	DNA encoding a Staphylococcus a
N_Geneseq_36:X20054	-	64.00	141.02	2.97	600	600	Enterococcus faecalis gene EF02
N_Geneseq_36:X59876	-	63.50	138.51	4.09	675	675	DNA encoding a Staphylococcus a
N_Geneseq_36:X07959	+	62.50	134.77	6.62	759	759	Flea serine protease gene SEQ I
N_Geneseq_36:X06815	+	65.00	131.79	9.69	1798	1798	Sequence encoding HBV core p
N_Geneseq_36:X74820	+	63.50	130.93	10.82	1373	1373	Firefly luciferase mRNA. New m
N_Geneseq_36:X798716	+	61.50	127.67	16.44	1168	1168	Staphylococcus aureus contig S
N_Geneseq_36:X20287	+	54.50	126.81	18.36	247	247	DNA encoding a S. pneumoniae p
N_Geneseq_36:X78691	+	55.50	125.19	22.60	363	363	Probe (188) for microbial genes
N_Geneseq_36:X90373	-	63.50	124.80	23.77	2438	2438	Staphylococcus aureus contig SE
N_Geneseq_36:X78146	-	55.50	124.35	25.81	400	400	DNA encoding amphiregulin pred
N_Geneseq_36:X74854	-	62.00	122.93	30.18	2045	2045	Staphylococcus aureus contig SE
N_Geneseq_36:X16270	+	54.00	122.86	30.46	318	318	Human fatty acid amide hydrola
N_Geneseq_36:X065373	+	54.50	122.78	30.77	360	360	0.3 kb recA fragment from pCEM
N_Geneseq_36:X780720	+	54.50	121.06	38.37	423	423	Fish gonadotropic hormone alpha
N_Geneseq_36:X63622	+	54.00	120.93	39.02	381	381	Type II topoisomerase database
N_Geneseq_36:X90869	+	51.50	120.00	43.99	232	232	Human RNase gene sequence. Cyt
N_Geneseq_36:X52067	-	58.00	119.78	45.24	1080	1080	Nucleotide sequence of clone Y2
N_Geneseq_36:X727047	-	58.00	119.64	46.05	1094	1094	Helicobacter polyptide GHPO
N_Geneseq_36:X13231	+	61.00	118.91	50.60	2653	2653	STR3 (suppressor of telomeric
N_Geneseq_36:X744331	+	64.00	118.25	48.44	4607	4607	Enterococcus faecalis genome c
N_Geneseq_36:X19206	+	61.50	118.89	50.67	2656	2656	B. burgdorferi decorin binding
N_Geneseq_36:X07456	-	58.50	118.75	51.60	1336	1336	B. burgdorferi dppA/dbpB genes
N_Geneseq_36:X07456	-	58.50	118.63	52.45	1203	1203	Homo sapiens HLIN-3 gene. Huma
N_Geneseq_36:X77449	+	50.50	118.47	53.53	212	212	Dengue virus type 4 envelope c
N_Geneseq_36:X25014	+	56.00	117.55	60.20	834	834	Staphylococcus aureus contig SE
N_Geneseq_36:X1417	-	61.00	117.41	61.31	2716	2716	H. pylori cell envelope transp
N_Geneseq_36:X50521	+	51.00	117.25	62.58	267	267	H. pylori GHPO 886 gene. New
N_Geneseq_36:X63623	+	54.00	117.21	62.92	540	540	Sequence encoding salmon gonad
N_Geneseq_36:X91174	+	52.00	117.16	63.29	340	340	Human EGF-human RNase sequence
N_Geneseq_36:X67236	+	59.00	117.09	63.90	1755	1755	Sequence encoding fish gonadotr
N_Geneseq_36:X065374	+	52.00	117.00	64.57	345	345	Pasteurella haemolytica trans
N_Geneseq_36:X24827	+	56.00	116.99	64.59	879	879	Fish gonadotropic hormone alpha
N_Geneseq_36:X63917	-	56.00	116.87	65.70	889	889	H. pylori ORF 02cei0809orf6. He
N_Geneseq_36:X63917	-	56.00	116.87	65.70	889	889	Mycobacterium tuberculosis ant

N\_Geneseq\_36:X00700 - 57.00 116.66 67.49 1145 ! Human secreted protein gene  
N\_Geneseq\_36:X30558 + 57.50 116.46 69.24 1311 ! H. pylori outer membrane pr  
N\_Geneseq\_36:X27327 + 57.50 115.73 76.05 1404 ! Streptococcus pneumoniae SP

seq\_name: N\_Geneseq\_36:X35210

seq\_documentation\_block:

ID X35210 standard; DNA; 1767 BP.

AC X35210;

AT 01-JUL-1999 (first entry)

DE Nucleotide sequence of PCV isolate Impl011-48121.

KW PCV isolate; type II porcine circovirus; PCV; PMWS;

KW porcine multisystemic wasting syndrome; pig; vaccine; ss.

OS porcine circovirus.

PN FR2769322-AL.

PD 09-APR-1999.

PF 22-JAN-1998; 000873.

PR 03-OCT-1997; FR-012382.

PA (MERI-) MERIAL SAS.

PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,

PI Haines D, Harding J, Hassard L, Meehan B;

DR WPI; 99-246948/21.

PT New type II porcine circovirus, used for, e.g. passive immunization

PS of pregnant sows

PS Claim 14; Fig 1; 48pp; French.

CC The present sequence represents the nucleotide sequence of PCV isolate

CC Impl011-48121. The specification describes a preparation of type II

CC porcine circovirus (PCV), which is particularly isolated from a lesion,

CC from a pig with symptoms of PMWS (porcine multisystemic wasting

CC syndrome). PCV (attenuated or inactivated), polypeptides derived from

CC it, and vectors that express these polypeptides are all useful in

CC vaccines, suitable for administration to adult or young pigs, or to

CC pregnant sows (for passive immunization of their offspring). DNA

CC isolated from PCV is used for in vivo or in vitro expression of viral

CC polypeptides, also as probes or primers for diagnosis in usual

CC hybridization or amplification assays. These polypeptides may also be

CC used diagnostically to detect PCV-specific antibodies, while antibodies

CC raised against the polypeptides can be used to detect antigens, in any

CC usual immunoassay format.

CC Sequence 1767 BP; 447 A; 360 C; 502 G; 458 T;

alignment\_scores:

Quality: 295.00 Length: 59

Ratio: 5.175 Gaps: 0

Percent Similarity: 96.610 Percent Identity: 94.915

alignment\_block:

US-09-209-961-7 x X35210/rev ..

Align seg 1/1 to reverse of: X35210 from: 1 to: 1767

1 MetLysCysThrLeuValPheGlnSerArgPheCysTlePheProLeuTh 17

||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

912 AGACGTGTACATAGTCTTCCAAATCAGCTTCTGCAATTTCCCGCTCAC 863

||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

17 rPheLysSerSerAlaSerProArgLysPheLeuThrAsnValThrGlyC 34

||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

862 TTTCAAAGTTCAGCCAGCCCGCGGAAATTTCTGCAACAGTTACAGGGT 813

||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

34 yscCysPheLaThrValThrArgIleProLeuSerAsnLysValLeuThr 50

||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

812 GCTGCTCTGCAAGGTCACAGAGCTCCGCTCTCCCAACAGGTACTCACA 763

||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

51 AlaValAspArgSerLeuArgCysPro 59

||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

762 GCAGTAGACAGGTCCTCGTGTCTCC 736

seq\_name: N\_Geneseq\_36:X35211

seq\_documentation\_block:

ID X35211 standard; DNA; 1767 BP.

AC X35211;

DT 01-JUL-1999 (first entry)  
 DE Nucleotide sequence of PCV isolate Imp1011-48285.  
 KW PCV isolate; type II porcine circovirus; PCV; PMWS;  
 KW porcine multisystemic wasting syndrome; pig; vaccine; ss.  
 OS Porcine circovirus.  
 PN FR2769322-Al.  
 PD 09-APR-1999.  
 PR 22-JAN-1998; 000873.  
 PR 03-OCT-1997; FR-012382.  
 PA (MERI-) Merial SAS.  
 PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,  
 PI Haines D, Harding J, Hassard L, Meehan B;  
 DR WPI; 99-246948/21.  
 PT New type II porcine circovirus, used for, e.g. passive immunization  
 of pregnant sows  
 PS Claim 14; Fig 2; 48pp; French.  
 CC The present sequence represents the nucleotide sequence of PCV isolate  
 CC Imp1011-48285. The specification describes a preparation of type II  
 CC porcine circovirus (PCV), which is particularly isolated from a lesion,  
 CC from a pig with symptoms of PMWS (porcine multisystemic wasting  
 CC syndrome). PCV (attenuated or inactivated), polypeptides derived from  
 CC it, and vectors that express these polypeptides are all useful in  
 CC vaccines, suitable for administration to adult or young pigs, or to  
 CC pregnant sows (for passive immunization of their offspring). DNA  
 CC isolated from PCV is used for in vivo or in vitro expression of viral  
 CC polypeptides, also as probes or primers for diagnosis in usual  
 CC hybridization or amplification assays. These polypeptides may also be  
 CC used diagnostically to detect PCV-specific antibodies, while antibodies  
 CC raised against the polypeptides can be used to detect antigens, in any  
 CC usual immunoassay format.  
 SQ Sequence 1767 BP; 448 A; 359 C; 500 G; 460 T;

alignment\_scores:  
 Quality: 295.00 Length: 59  
 Ratio: 5.175 Gaps: 0  
 Percent similarity: 96.610 Percent Identity: 94.915

alignment\_block:  
 US-09-209-961-7 x X35211/rev ..  
 Align seg 1/1 to reverse of: X35211 from: 1 to: 1767  
 1 MetLysCysThrLeuValPheGlnSerArgPheCysIlePheProLeuThr 17  
 ||| ||||||||||||||||||||||||||||||||||||||||||||  
 912 ATGACGTGTACATAGTCTTCCATCAGCTTCGCAATTTCCGCTCAC 863  
 17 rPheLysSerSerAlaSerProArgLysPheLeuThrAsnValThrGlyC 34  
 ||| ||||||||||||||||||||||||||||||||||||||||||||  
 862 TTTCAAAGTTTCAGCCAGCCGCGGAAATTTCTGACAAACGTTACAGGT 813  
 34 yScysPheAlaThrValThrArgIleProLeuSerAsnLysValLeuThr 50  
 ||| ||||||||||||||||||||||||||||||||||||||||||||  
 812 GCTGCTCTGCAACGGTCCACGAGCTCCGCTCTCCAAACAGGTACTCACA 763  
 51 AlaValAspArgSerLeuArgCysPro 59  
 ||| ||||||||||||||||||||||||||||||||||||||||||||  
 762 GCAGTAGACAGGTCACTCCGTTGCTCC 736  
 seq\_name: N\_Geneseq\_36:X35012  
 seq\_documentation\_block:  
 ID X35012 standard; DNA; 1768 BP.  
 AC X35012;  
 DT 01-JUL-1999 (first entry)  
 DE Genomic DNA sequence of PCV strain 999PCV.  
 KW PCV; strain PCV999; pig; PMWS; porcine multisystemic wasting syndrome;  
 KW vaccine; ss.  
 OS Porcine circovirus  
 PN FR2769321-Al.  
 PD 09-APR-1999.  
 PR 03-OCT-1997; 012382.  
 PR 03-OCT-1997; FR-012382.

PA (MERI-) Merial SAS.  
 PA (UYBE-) UNIV QUEENS BELFAST.  
 PA (UYSA-) UNIV SASKATCHEWAN  
 PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,  
 PI Haines D, Harding J, Hassard L, Meehan B;  
 DR WPI; 99-246947/21.  
 PT New porcine circovirus from animals with porcine systemic wasting  
 syndrome  
 PS Claim 13; Fig 1; 35pp; French.  
 CC The present sequence represents the genomic sequence of porcine  
 CC circovirus (PCV) strain PCV999. The sequence is obtained from PCV  
 CC isolated from a physiological or tissue sample, particularly from  
 CC a lesion, from a pig showing symptoms of PMWS (porcine multisystemic  
 CC wasting syndrome), or cultured cells, infected with PCV isolated from  
 CC such pigs. PCV (attenuated or inactivated), polypeptides derived from  
 CC it, and vectors that express these polypeptides are all useful in  
 CC vaccines, suitable for administration to adult or young pigs, or to  
 CC pregnant sows (for passive immunization of their offspring). DNA  
 CC isolated from PCV is used for in vivo or in vitro expression of viral  
 CC polypeptides, also as probes or primers for diagnosis in usual  
 CC hybridization or amplification assays. These polypeptides may also be  
 CC used diagnostically to detect PCV-specific antibodies, while antibodies  
 CC raised against the polypeptides can be used to detect antigens, in any  
 CC usual immunoassay format.  
 SQ Sequence 1768 BP; 453 A; 359 C; 486 G; 455 T;

alignment\_scores:  
 Quality: 295.00 Length: 59  
 Ratio: 5.175 Gaps: 0  
 Percent similarity: 96.610 Percent Identity: 94.915

alignment\_block:  
 US-09-209-961-7 x X35012/rev ..  
 Align seg 1/1 to reverse of: X35012 from: 1 to: 1768  
 1 MetLysCysThrLeuValPheGlnSerArgPheCysIlePheProLeuThr 17  
 ||| ||||||||||||||||||||||||||||||||||||||||||||  
 913 ATGACGTGTACATAGTCTTCCATCAGCTTCGCAATTTCCGCTCAC 864  
 17 rPheLysSerSerAlaSerProArgLysPheLeuThrAsnValThrGlyC 34  
 ||| ||||||||||||||||||||||||||||||||||||||||||||  
 863 TTTCAAAGTTTCAGCCAGCCGCGGAAATTTCTGACAAACGTTACAGGT 814  
 34 yScysPheAlaThrValThrArgIleProLeuSerAsnLysValLeuThr 50  
 ||| ||||||||||||||||||||||||||||||||||||||||||||  
 813 GCTGCTCTGCAACGGTCCACGAGCTCCGCTCTCCAAACAGGTACTCACA 764  
 51 AlaValAspArgSerLeuArgCysPro 59  
 ||| ||||||||||||||||||||||||||||||||||||||||||||  
 763 GCAGTAGACAGGTCACTCCGTTGCTCT 737  
 seq\_name: N\_Geneseq\_36:X35213  
 seq\_documentation\_block:  
 ID X35213 standard; DNA; 1768 BP.  
 AC X35213;  
 DT 01-JUL-1999 (first entry)  
 DE Nucleotide sequence of PCV isolate Imp1010.  
 KW PCV isolate; type II porcine circovirus; PCV; PMWS;  
 KW porcine multisystemic wasting syndrome; pig; vaccine; ss.  
 OS Porcine circovirus.  
 PN FR2769322-Al.  
 PD 09-APR-1999.  
 PR 22-JAN-1998; 000873.  
 PR 03-OCT-1997; FR-012382.  
 PA (MERI-) Merial SAS.  
 PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,  
 PI Haines D, Harding J, Hassard L, Meehan B;  
 DR WPI; 99-246948/21.  
 PT New type II porcine circovirus, used for, e.g. passive immunization  
 of pregnant sows

```
PS Claim 14; Fig 4: 48pp; French.
CC The present sequence represents the nucleotide sequence of PCV isolate
CC Impl010. The specification describes a preparation of type II
CC porcine circovirus (PCV), which is particularly isolated from a lesion,
CC from a pig with symptoms of PMWS (porcine multisystemic wasting
CC syndrome). PCV (attenuated or inactivated), polypeptides derived from
CC it, and vectors that express these polypeptides are all useful in
CC vaccines, suitable for administration to adult or young pigs, or to
CC pregnant sows (for passive immunization of their offspring). DNA
CC isolated from PCV is used for in vivo or in vitro expression of viral
CC polypeptides, also as probes or primers for diagnosis in usual
CC hybridization or amplification assays. These polypeptides may also be
CC used diagnostically to detect PCV-specific antibodies, while antibodies
CC raised against the polypeptides can be used to detect antigens, in any
CC usual immunoassay format.
SQ Sequence 1768 BP; 451 A; 361 C; 495 G; 461 T;

alignment_scores:
Quality: 295.00 Length: 59
Ratio: 5.175 Gaps: 0
Percent Similarity: 96.610 Percent Identity: 94.915

alignment_block:
US-09-209-961-7 x X35213/rev ..

Align seg 1/1 to reverse of: X35213 from: 1 to: 1768

1 MetLysCysThrLeuValPheGlnSerArgPheCysIlePheProLeuTh 17
|||||
912 ATGACGTGTACATTGGTCTTCCAATCAGCTTCTGCAATTTCCCGCTCAC 863

17 rPheLysSerSerAlaSerProArgLysPheLeuThrAsnValThrGlyC 34
|||||
862 TTTCAAAAGTTTCAGCCAGCCCGCGGAAATTTCTGACAAACGTTACAGG 813

34 yScysPheAlaThrValThrArgIleProLeuSerAsnLysValLeuThr 50
|||||
812 GCTGCTCTGCACAGGTACACAGCTCCGCTCTCCAAACAGGTACTCACA 763

51 AlaValAspArgSerLeuArgCysPro 59
|||||
762 GCAGTAGACAGGTCACTCGTGTCTCT 736

seq_name: N_Geneseq_36:X35212

seq_documentation_block:
ID X35212 standard; DNA; 1768 BP.
AC X35212;
DT 01-JUL-1999 (first entry)
DE Nucleotide sequence of PCV isolate Imp999 (corrected version).
KW PCV isolate; type II porcine circovirus; PCV; PMWS;
KW porcine multisystemic wasting syndrome; pig; vaccine; ss.
OS Porcine circovirus.
PN FR2769322-Al.
PD 09-APR-1999.
PF 22-JAN-1998; 000873.
PR 03-OCT-1997; FR-012382.
PA (MERI-) MERIAL SAS.
PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,
PI Haines D, Harding J, Hassard L, Meehan B;
DR WPI: 99-246948/21.
PT New type II porcine circovirus, used for, e.g. passive immunization
PT of pregnant sows
PS Claim 14; Fig 3: 48pp; French.
CC The present sequence represents the nucleotide sequence of PCV isolate
CC Imp999. The specification describes a preparation of type II
CC porcine circovirus (PCV), which is particularly isolated from a lesion,
CC from a pig with symptoms of PMWS (porcine multisystemic wasting
CC syndrome). PCV (attenuated or inactivated), polypeptides derived from
CC it, and vectors that express these polypeptides are all useful in
CC vaccines, suitable for administration to adult or young pigs, or to
CC pregnant sows (for passive immunization of their offspring). DNA
SQ Sequence 1768 BP; 451 A; 361 C; 495 G; 461 T;
```

```
CC isolated from PCV is used for in vivo or in vitro expression of viral
CC polypeptides, also as probes or primers for diagnosis in usual
CC hybridization or amplification assays. These polypeptides may also be
CC used diagnostically to detect PCV-specific antibodies, while antibodies
CC raised against the polypeptides can be used to detect antigens, in any
CC usual immunoassay format.
SQ Sequence 1768 BP; 453 A; 360 C; 494 G; 461 T;

alignment_scores:
Quality: 295.00 Length: 59
Ratio: 5.175 Gaps: 0
Percent Similarity: 96.610 Percent Identity: 94.915

alignment_block:
US-09-209-961-7 x X35212/rev ..

Align seg 1/1 to reverse of: X35212 from: 1 to: 1768

1 MetLysCysThrLeuValPheGlnSerArgPheCysIlePheProLeuTh 17
|||||
912 ATGACGTGTACATTGGTCTTCCAATCAGCTTCTGCAATTTCCCGCTCAC 863

17 rPheLysSerSerAlaSerProArgLysPheLeuThrAsnValThrGlyC 34
|||||
862 TTTCAAAAGTTTCAGCCAGCCCGCGGAAATTTCTGACAAACGTTACAGG 813

34 yScysPheAlaThrValThrArgIleProLeuSerAsnLysValLeuThr 50
|||||
812 GCTGCTCTGCACAGGTACACAGCTCCGCTCTCCAAACAGGTACTCACA 763

51 AlaValAspArgSerLeuArgCysPro 59
|||||
762 GCAGTAGACAGGTCACTCGTGTCTCT 736

seq_name: N_Geneseq_36:X35214

seq_documentation_block:
ID X35214 standard; DNA; 1768 BP.
AC X35214;
DT 01-JUL-1999 (first entry)
DE Nucleotide sequence of PCV isolate Imp999 (uncorrected version).
KW PCV isolate; type II porcine circovirus; PCV; PMWS;
KW porcine multisystemic wasting syndrome; pig; vaccine; ss.
OS Porcine circovirus.
PN FR2769322-Al.
PD 09-APR-1999.
PF 22-JAN-1998; 000873.
PR 03-OCT-1997; FR-012382.
PA (MERI-) MERIAL SAS.
PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,
PI Haines D, Harding J, Hassard L, Meehan B;
DR WPI: 99-246948/21.
PT New type II porcine circovirus, used for, e.g. passive immunization
PT of pregnant sows
PS Claim 14; Fig 6; 48pp; French.
CC The present sequence represents the nucleotide sequence of PCV isolate
CC Imp999. The specification describes a preparation of type II
CC porcine circovirus (PCV), which is particularly isolated from a lesion,
CC from a pig with symptoms of PMWS (porcine multisystemic wasting
CC syndrome). PCV (attenuated or inactivated), polypeptides derived from
CC it, and vectors that express these polypeptides are all useful in
CC vaccines, suitable for administration to adult or young pigs, or to
CC pregnant sows (for passive immunization of their offspring). DNA
CC isolated from PCV is used for in vivo or in vitro expression of viral
CC polypeptides, also as probes or primers for diagnosis in usual
CC hybridization or amplification assays. These polypeptides may also be
CC used diagnostically to detect PCV-specific antibodies, while antibodies
CC raised against the polypeptides can be used to detect antigens, in any
CC usual immunoassay format.
SQ Sequence 1768 BP; 453 A; 359 C; 486 G; 455 T;
```

alignment\_scores:  
 Quality: 295.00 Length: 59  
 Ratio: 5.175 Gaps: 0  
 Percent Similarity: 96.610 Percent Identity: 94.915

alignment\_block:  
 US-09-209-961-7 x X35214/rev ..

Align seg 1/1 to reverse of: X35214 from: 1 to: 1768

```

1 MetLysCysThrLeuValPheGlnSerArgPheCysIlePheProLeuTh 17
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
913 ATGACGTGTACATTGGTCTTCCAATCAGCTTCGCAATTTCCCGCTCAC 864

17 rPheLysSerSerAlaSerProArgLysPheLeuThrAsnValThrGlyC 34
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
863 TTTCAAAAGTTTCAGCCAGCCGCGGAAATTTCTGACAAAACGTTACAGGT 814

34 yScysPheAlaThrValThrArgIleProLeuSerAsnLysValLeuThr 50
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
813 GCTGCTCTGCAACGGTCACAGACTCCCGCTCTCCAAACAGGTACTCACA 764

```

51 AlavalaspargSerLeuArgCysPro 59

763 GCAGTAGACAGGTCACCTCGTTGTCT 737

seq\_name: N\_Geneseq\_36:X35013

seq\_documentation\_block:

ID X35013 standard; DNA; 1759 BP.  
 AC X35013;  
 DT 01-JUL-1999 (first entry)  
 DE Nucleotide sequence which has homology to PCV sequence.  
 KW PCV; strain PCV999; pig; PMWS; porcine multisystemic wasting syndrome;  
 KW vaccine; ss.  
 OS Sus sp.  
 PN FR2769321-Al.  
 PD 09-APR-1999.  
 PF 03-OCT-1997; 012382.  
 PR 03-OCT-1997; FR-012382.  
 PA (MERI-) MERIAL SAS.  
 PA (UYBE-) UNIV QUEENS BELFAST.  
 PA (UYGA-) UNIV SASKATCHEWAN.  
 PI Allan G, Chapuis GE, Charreyre CE, Clark E, Ellis J,  
 PI Haines D, Harding J, Hassard L, Meehan B;  
 DR WPI; 99-245947/21.  
 PT New porcine circovirus from animals with porcine systemic wasting  
 PT syndrome  
 PS Disclosure; Fig 2; 35pp; French.

CC The specification describes a genomic sequence of porcine  
 CC circovirus (PCV) strain PCV999. The sequence is obtained from PCV  
 CC isolated from a physiological or tissue sample, particularly from  
 CC a lesion, from a pig showing symptoms of PMWS (porcine multisystemic  
 CC wasting syndrome), or cultured cells, infected with PCV isolated from  
 CC such pigs. PCV (attenuated or inactivated), polypeptides derived from  
 CC it, and vectors that express these polypeptides are all useful in  
 CC vaccines, suitable for administration to adult or young pigs, or to  
 CC pregnant sows (for passive immunization of their offspring). DNA  
 CC isolated from PCV is used for in vivo or in vitro expression of viral  
 CC polypeptides, also as probes or primers for diagnosis in usual  
 CC hybridization or amplification assays. These polypeptides may also be  
 CC used diagnostically to detect PCV-specific antibodies, while antibodies  
 CC raised against the polypeptides can be used to detect antigens, in any  
 CC usual immunoassay format. The present sequence appears in the  
 CC specification.  
 SQ Sequence 1759 BP; 475 A; 510 C; 342 G; 432 T;

alignment\_scores:  
 Quality: 231.00 Length: 59  
 Ratio: 4.529 Gaps: 0  
 Percent Similarity: 86.441 Percent Identity: 79.661

alignment\_block:

US-09-209-961-7 x X35013 ..

Align seg 1/1 to: X35013 from: 1 to: 1759

```

1 MetLysCysThrLeuValPheGlnSerArgPheCysIlePheProLeuTh 17
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
860 ATGACGTGTACAGCTGTCTTCCAATCAGCTGTGCATCTTCCCGCTCAC 909

17 rPheLysSerSerAlaSerProArgLysPheLeuThrAsnValThrGlyC 34
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
910 TTTCAAAAGTTTCAGCCAGCCGCGGAAATTTCTCATACTACAGGGA 959

34 yScysPheAlaThrValThrArgIleProLeuSerAsnLysValLeuThr 50
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
960 ACTGTCGCTACAGTACCAAGACCCGCTCTCCAAAAGGTACTCACA 1009

51 AlavalaspargSerLeuArgCysPro 59
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1010 GCAGTAGACAGGTCGCTCGCTTCCCTCC 1036

```

seq\_name: N\_Geneseq\_36:T84052

seq\_documentation\_block:

ID T84052 standard; DNA; 925 BP.  
 AC T84052;  
 DT 27-AUG-1998 (first entry)  
 DE DNA encoding a Staphylococcus aureus protein of unknown function.  
 KW Staphylococcus aureus protein; ribozyme; antisense sequence; control;  
 KW Staphylococcal gene; regulatory element; bacterial gene expression;  
 KW vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;  
 KW toxic shock syndrome; ss.  
 OS Staphylococcus aureus  
 FH Key  
 FT CDS 285..452  
 FT /\*tag= a

WO9730070-Al.  
 21-AUG-1997.  
 PD 19-FEB-1997; U02318.  
 PF 20-FEB-1996; US-011888.  
 PR (SMIK ) SMITHKLINE BEECHAM CORP.  
 PI Black MT, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO,  
 PI Pratt JM, Reichard RW, Rosenberg M, Ward JM;  
 DR WPI; 97-424969/39.  
 DR P-PSDB; W28121  
 PT Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used  
 PT to isolate antimicrobial compounds, and in vaccines against S.  
 PT aureus infection  
 PS Claim 9; Page 846; 989pp; English.  
 CC The present sequence encodes a Staphylococcus aureus protein of  
 CC unknown function. The present sequence was isolated from a  
 CC library of clones of S. aureus WCUH 29 in Escherichia coli. The DNA  
 CC sequence can be used in the construction of ribozymes and antisense  
 CC sequences to control the expression of Staphylococcal genes. The DNA  
 CC sequence is also useful as a source of regulatory elements for the  
 CC control of bacterial gene expression. The encoded protein may be used  
 CC to produce vaccines to enable a host to produce specific antibodies  
 CC with antibacterial action. These vaccines and antibodies would protect  
 CC a host against infection by S. aureus, and conditions relating to  
 CC Staphylococcal infection, e.g. Staphylococcal food poisoning, scaled  
 CC skin syndrome, and toxic shock syndrome.  
 SQ Sequence 925 BP; 327 A; 135 C; 198 G; 261 T;

alignment\_scores:  
 Quality: 69.50 Length: 44  
 Ratio: 2.574 Gaps: 1  
 Percent Similarity: 61.364 Percent Identity: 38.636

alignment\_block:

US-09-209-961-7 x T84052/rev ..

Align seg 1/1 to reverse of: T84052 from: 1 to: 925



```

FH Key      Location/Qualifiers
FT CDS      1..741
FT FT       /*tag= a
FT FT       /product= "serine protease"
FT FT       /transl_except= (pos:721..722,aa:Xaa)
FT FT       /note= "Xaa= stop codon"

PN WO9740058-A1.
PD 30-OCT-1997.
PF 24-APR-1997; U06121.
PR 04-APR-1997; US-042945.
PR 24-APR-1996; US-639075.
PR 15-NOV-1996; US-749699.
PA (HESK-) HESKA CORP.
PI Frank GR, Gaines PJ, Grievle RB, Hunter SW, Rushlow KE,
PI Silver G, Stiegler GL;
DR WPI: 98-076762/07.
DR P-PSDB: W41974.
PT New flea protease genes and proteins - used in vaccine compositions
PT for the prophylaxis and treatment of flea infestation, especially in
PT cats or dogs
PS Claim 1: Page 259-260; 318pp; English.
CC The present sequence encodes a novel flea serine protease. The
CC protease, its mimetopes, antibodies (Ab) and inhibitors of the
CC protein, as well as the DNA encoding the protein, may all be used in
CC therapeutic compositions to reduce flea protease activity (especially
CC immunoglobulin protease) and so reduce flea infestation, especially in
CC cats or dogs. Alternatively, flea larvae may ingest the faeces of adult
CC fleas which comprises anti-protease antibodies produced by a host animal
CC in response to administration of the protein. Therapeutic compositions
CC may further comprise a compound that reduces haematophagous ectoparasite
CC burden by a method other than by reducing flea immunoglobulin protease
CC activity. The novel flea DNA encoding the protein can also be used to
CC produce recombinant protein, and fragments of it are used as probes and
CC primers for identification and isolation of related sequences, also as
CC antisense, triplex-forming agents and ribozymes for inhibition of the
CC synthesis of the protein. Ab are also useful for screening expression
CC libraries, to purify the protein and to target cytotoxins to fleas.
SQ Sequence 759 BP; 255 A; 123 C; 172 G; 209 T;

alignment_scores:
  Quality: 62.50      Length: 56
  Ratio: 1.894       Gaps: 2
  Percent Similarity: 58.929      Percent Identity: 30.357

alignment_block:
US-09-209-961-7 x V04597/rev ..
Align seg 1/1 to reverse of: V04597 from: 1 to: 759

11 PheCysIlePheProLeuThr.....PheLysSerSerAlaSerPr 24
359 TTITGACATTTCGCTGAATTTATTTATTTATTCGGAAGTTTACAGTGC 310

24 OArgLysPheLeu.....ThrAsnValThrGlyCysCysPheA 37
309 CAATATCATTTGGCGCAGTTGTTTCATCAACTGTCAGTATAGTATTTT 260

37 laThrValThrArgIleProLeuSerAsnLysValLeuThrAlaValAsp 53
259 CTATCTATATCGTTCCACCTACTCTACCTACAGTTTATTCACGCGGCAACG 210

54 ArgSerLeuArgCysPro 59
209 ACTTCAAAACGTTGTCC 192

seq_name: N_Geneseq_36:Q67959

seq_documentation_block:
ID Q67959 standard; DNA; 627 BP.
AC Q67959;
DT 02-JAN-1995 (first entry)
DE Sequence encoding HBV core-pres1.

```

```

KW Core antigen; recombinant replicable vaccinia virus; hepatitis;
KW prevention; therapy; epitope; ss.
OS Hepatitis B virus
FH Key      Location/Qualifiers
FT CDS      1..627
FT FT       /*tag= a

PN WO9412617-A.
PD 09-JUN-1994.
PF 24-NOV-1993; U11474.
PR 25-NOV-1992; US-982211.
PA (ITBI-) INT BIOTECHNOLOGY LAB INC.
PI Bernstine EG, Lewis T, Okeefe RW, Souw PTS;
DR WPI: 94-200247/24.
DR P-PSDB: R55279.
PT Prevention and treatment of hepatitis - using recombinant
PT replicable vaccinia viruses contg. hepatitis B virus surface and
PT core antigen nucleotide sequences
PS Example: Fig 44; 252pp; English.
CC A vector was constructed to allow expression of a core-pres1 fusion
CC polypeptide (called core-pres1* because only a portion of pres1 is
CC terminus of the fusion protein, of AA residues 1-145 or core, a 3 AA
CC spacer (SAC), AAs 1-56 of pres1, and a 4 AA tail (RPTS). To make
CC this fusion vector, a piece of the pres1 region was amplified by PCR
CC from pUEH-04. One primer used was Q67983 which hybridises from bps
CC -22 to -3 relative to the ATG of pres1 and is designed to introduce
CC a BspEI site and to destroy a potential termination codon. A second
CC primer was Q67984 which hybridises to the opposite strand from bps
CC +171 to +152 of the pres1 region and is designed to create an
CC EcoRI site upon amplification. The resulting plasmid, pHTL-26,
CC consists of the modified p7.5 promoter oriented so as to drive
CC expression of the core-pres1* fusion. The sequences are given in
CC Q67959 and R55279.
SQ Sequence 627 BP; 154 A; 157 C; 138 G; 178 T;

alignment_scores:
  Quality: 61.50      Length: 54
  Ratio: 1.809       Gaps: 3
  Percent Similarity: 52.963      Percent Identity: 35.185

alignment_block:
US-09-209-961-7 x Q67959 ..
Align seg 1/1 to: Q67959 from: 1 to: 627

9 SerArgPheCysIlePheProLeuThrPheLysSerSerAlaSerProA 25
47 TCTCGTTTTCCTCTCTGACTTCTTCTTCCTCGT...CAGAGATCTCTCTA 93

25 xgLySPheLeuThrAsnValThrGlyCys.....Cys 35
94 GACACCCCTCAGCTCTGTATCGGGAAGCCTTAGAGTCTCTCTGAGGATGC 143

36 PheAlaThrValThrArgIleProLeuSerAsnLysValLeuThrAlaVa 52
144 TTACCTTACACACACCCACTCAGGCAAGCCATTCCTCTGCTGGGGGGAAT 193

52 lAspArgSer 55
194 TGATGACTCT 203

seq_name: N_Geneseq_36:X06815

seq_documentation_block:
ID X06815 standard; mRNA; 1798 BP.
AC X06815;
DT 26-APR-1999 (first entry)
DE Firefly luciferase mRNA.
KW Gene activation; gene expression; cancer; antisense; therapy;
KW beta-galactosidase; luciferase; firefly; ss.
OS Photinus sp.
PN WO9858944-A1.

```

PD 30-DEC-1998.  
PF 24-JUN-1998; UI3093.  
PR 25-JUN-1997; US-050772.  
PA (BLAC/) BLACK C A.  
PI BLACK CA;  
DR WPI: 99-095360/08.  
PT New masked targeted expression cassette - useful for regulating  
PT gene expression, and preventing neoplastic cell growth  
PS Disclosure: Page 28-29; 43pp; English.  
CC This is a firefly luciferase full-length mRNA sequence. A new  
CC method for activating genes of interest uses a masked targeted  
CC expression cassette which expresses a gene product only in the  
CC presence of a target molecule. Plasmid pCI-Neo was designed for  
CC the production of the sense RNA strand of a targeted expression  
CC cassette. Alternative flanking sequences (see X06801-07)  
CC corresponding to portions of the firefly luciferase mRNA can be  
CC inserted into the multiple cloning sites (see X06800) of the  
CC plasmid such that transcription from the T7 promoter yields RNA  
CC comprising luciferase segment-Kozak-beta-galactosidase. Sense  
CC strand RNA of a masked targeted cassette is produced by in vitro  
CC transcription of the construct. Antisense sequences (see X06808-14)  
CC corresponding to portions of the target molecule are hybridised to  
CC complementary flanking sequences of the sense strand of the  
CC targeted cassette. The hybridised mixture is introduced to an in  
CC vitro translation mixture containing ribosomes and full-length  
CC firefly luciferase RNA. After completion of translation, the  
CC mixture is assayed for beta-galactosidase and luciferase  
CC activities. Negative luciferase and positive beta-galactosidase  
CC activity indicates successful inhibition of the target molecule and  
CC successful expression of the gene of interest. The cassette is  
CC useful for the treatment of disease and for preventing the  
CC proliferation of neoplastic cells. The cassette only releases the  
CC antisense strand and expresses a gene product in the presence of a  
CC target molecule, so is more potent, reliable, and specific than  
CC prior art antisense technology.  
SQ Sequence 1798 BP; 540 A; 375 C; 423 G; 460 U;

alignment\_scores:  
Quality: 65.00 Length: 74  
Ratio: 1.912 Gaps: 6  
Percent Similarity: 45.946 Percent Identity: 33.784

alignment\_block:  
US-09-209-961-7 x X06815/rev ..

Align seg 1/1 to reverse of: X06815 from: 1 to: 1798

11 PheCysIlePheProLeuThrPheLysSerAlaSerProArgLysPh 27  
||||| ||||||| ||||||| ||||||| ||||||| ||  
471 TTTTGC.....ACGTCAAAATTTTTCGAACCC.....TT 440

27 eLeu...ThrAsnValThr...GlyCys..... 34  
||||| ||||||| ||||||| ||||||| ||||||| ||  
439 TTTCGAACAACACTACGGTAGGTGGCAATGTTCTACTGTTGACCA 390

35 .....CysPheAlaThrValThrArgIleProLeu 44  
||||| ||||||| ||||||| ||||||| ||||||| ||  
389 ATTACAGTTCATTATAAATGTCTTCGGGGGCAACTGCAACTCCGATA 340

45 SerAsn.....LysValLeuThrAl 51  
||||| ||||||| ||||||| ||||||| ||||||| ||  
339 AATACGCGGCCCAACACCGGCATAAAGATTGAAGAGAGTTTTCATGTC 290

51 aValAspArgSerLeuArgCys 58  
||||| ||||||| ||||||| ||||||| ||||||| ||  
289 ATACGACGATCTGTGATTGT 268

seq\_name: N\_Geneseq\_36:V74820

seq\_documentation\_block:  
ID V74820 standard; DNA; 1373 BP.  
AC V74820;

DT 16-MAR-1999 (first entry)  
DE Staphylococcus aureus contig SEQ ID #509.  
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;  
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
KW skin infection; surgical wound infection; scalded skin syndrome;  
KW toxic shock syndrome; ds.  
OS Staphylococcus aureus.  
FH Key Location/Qualifiers  
FT misc\_feature 961..1020  
FT /\*tag=a  
FT /note= "these bases represent a line of missing text in  
FT the sequence listing in the specification. They  
FT are included to maintain the nucleotide numbering  
FT given in the specification for this DNA sequence"  
PN EP-786519-A2.  
PD 30-JUL-1997. 100117.  
PF 07-JAN-1997; US-009861.  
PR 05-JAN-1996; US-009861.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,  
PI Rosen CA;  
DR WPI: 97-374922/35.  
PT Polynucleotide(s) and proteins derived from Staphylococcus aureus  
PT stored on computer readable medium and used in the production of  
PT anti-S.aureus vaccines  
PS Claim 1: Page 1436-1437; 3271pp; English.  
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences  
CC of the invention. The DNA sequences are recorded on a computer readable  
CC medium, preferably selected from a floppy or hard disk, random access  
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
CC the S.aureus DNA sequences allows putative functions to be assigned so  
CC that protein-encoding or regulatory regions of commercial, therapeutic or  
CC industrial importance can be obtained. Specifically, sequences which are  
CC likely to encode antigens have been identified and these polypeptides can  
CC be used in a vaccine composition against S.aureus infection. The  
CC polypeptides can also be used in a kit for the immunodetection of  
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,  
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
CC skin and surgical wound infections, scalded skin syndrome, toxic shock  
CC syndrome, etc. Organisms transformed with the DNA sequences can be used  
CC for recombinant production of the polypeptides. The new DNA sequences  
CC (and their fragments) are useful as primers or probes for isolating  
CC homologues of any of the S.aureus DNA sequences contained on the  
CC computer readable medium.  
SQ Sequence 1373 BP; 363 A; 258 C; 183 G; 506 T;

alignment\_scores:  
Quality: 63.50 Length: 32  
Ratio: 3.175 Gaps: 1  
Percent Similarity: 62.500 Percent Identity: 46.875

alignment\_block:  
US-09-209-961-7 x V74820 ..

Align seg 1/1 to: V74820 from: 1 to: 1373

15 ProLeuThrPheLysSerAlaSerProArgLysPheLeuThrAsnVa 31  
||||| ||||||| ||||||| ||||||| ||||||| ||  
505 CCATATATCTTTTAGTTCATCGATAGTTTGTAGCTAAATTTGCTACTCGAAT 554

31 lhrGlyCysCys.....PheAlaThrVal 39  
||||| |||  
555 AACTGGTACATGTTCAATTGCACCTGTTGAGGCTTTTGCAACTGTT 600

seq\_name: N\_Geneseq\_36:T98716

seq\_documentation\_block:  
ID T98716 standard; DNA; 1168 BP.  
AC T98716;  
DT 09-NOV-1998 (first entry)  
DE DNA encoding a S. pneumoniae protein of unknown function.  
KW Streptococcus pneumoniae protein; genetic immunisation; antagonist;





OM of: US-09-209-961-7 to: Issued\_Patents\_NA:\* out\_format : pfs  
Date: Dec 31, 1999 12:26 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-1998 Compugen Ltd.

## Command line parameters:

-MODEL=frame1.p2n.model -DRV=xlp  
-Q/cgn2\_1/USPTO.spool/US09209961/runat.22121999\_101043\_17921/app\_query.fasta.1  
-DB=Issued\_Patents\_NA -OFMT=fastap -SUFFIX=rni -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blossum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=escore  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09209961  
-NCPU=6 -ICPU=3 -NO\_XLPXY -WAIT -THREADS=1

## Search information block:

Query: US-09-209-961-7  
Query length: 59  
Database: Issued\_Patents\_NA\*  
Database sequences: 195662  
Database length: 53930012  
Search time (sec): 88.010000

score_list:	Sequence	Strd Orig	ZScore	Escore	Len	Documentation
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-09-132-619-7	2330	1	62.50	121.79	21.84	
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-470-179-139	423	1	54.50	120.14	27.00	
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-125-462-8	381	1	54.00	120.01	27.44	
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-891-848-9	381	1	54.00	120.01	27.44	
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-09-132-619-1	3115	1	62.50	118.72	32.39	
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-431-080-17	1094	1	58.00	118.71	32.44	
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-431-080-30	1094	1	58.00	118.71	32.44	
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-938-534-17	1094	1	58.00	118.71	32.44	
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-938-534-30	1094	1	58.00	118.71	32.44	
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-09-132-619-5	3251	1	62.50	118.27	34.32	
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-589-711-1	2653	1	61.50	117.96	35.72	
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-739-485-6	1336	1	58.00	117.82	36.33	
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-125-462-9	540	1	54.00	116.32	44.04	
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-891-848-9	540	1	54.00	116.32	44.04	
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-009-261-4	492	1	53.50	116.08	45.45	
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-451-715A-7	2973	1	60.50	114.29	57.17	
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-867-129-1	2529	1	58.50	111.08	86.33	
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-791-213D-14	2147	1	57.50	110.35	94.81	
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-293-150A-14	210	1	47.00	109.08	111.56	
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-972-387-7	210	1	49.00	108.78	115.86	
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-431-412-7	344	1	49.00	108.78	115.86	
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-057-971-7	344	1	49.00	108.78	115.86	
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-743-518-25	520	1	50.50	108.11	126.36	
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-450-384-7	93	1	43.00	107.84	130.71	
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-309-604-1	342	1	48.50	107.61	134.61	
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-450-384-6	99	1	43.00	107.18	142.27	
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-392-546C-3	617	1	50.50	106.30	159.35	
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-093-144-6	444	1	49.00	106.08	163.77	
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-755-462-1	500	1	49.50	106.06	164.30	
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-199-317-1	500	1	49.50	106.06	164.30	
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-450-065-1	500	1	50.50	106.01	165.33	
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-450-595-1	634	1	50.50	106.01	165.33	
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-840-683-1	1192	1	53.00	105.49	176.76	
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-972-387-27	379	1	48.00	105.30	181.19	
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-431-412-27	379	1	48.00	105.30	181.19	
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-057-971-27	379	1	48.00	105.30	181.19	
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-068-395A-2	1755	1	54.50	105.09	186.01	
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-464-365-2	1755	1	54.50	105.09	186.01	
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-791-213D-88	313	1	47.00	104.86	191.68	
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-972-387-1	313	1	47.00	104.86	191.68	
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-431-412-1	313	1	47.00	104.86	191.68	

/cgn2\_6/ptodata/1/ina/5B\_COMB.seq:US-08-057-971-1 + 47.00 104.86 191.68 313  
/cgn2\_6/ptodata/1/ina/5C\_COMB.seq:US-08-293-150A-88 + 47.00 104.86 191.68 313  
/cgn2\_6/ptodata/1/ina/5A\_COMB.seq:US-08-370-975B-4 + 48.00 104.67 196.27 402

seq\_name: /cgn2\_6/ptodata/1/ina/5D\_COMB.seq:US-09-132-619-7

## seq\_documentation\_block:

; Sequence 7, Application US/09132619B  
; Patent No. 5958697  
; GENERAL INFORMATION:  
; APPLICANT: Shan, Bei  
; TITLE OF INVENTION: CYP7 Promoter-Binding Factors  
; FILE REFERENCE: T97-013  
; CURRENT APPLICATION NUMBER: US/09/132,619B  
; CURRENT FILING DATE: 1998-08-11  
; EARLIER APPLICATION NUMBER: 60/067,708  
; EARLIER FILING DATE: 1997-12-08  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 2330  
; TYPE: DNA  
; ORGANISM: human  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (363)..(1862)  
US-09-132-619-7

## alignment\_scores:

Quality: 62.50 Length: 32  
Ratio: 2.717 Gaps: 1  
Percent Similarity: 71.875 Percent Identity: 43.750

## alignment\_block:

US-09-209-961-7 x US-09-132-619-7

Align seg 1/1 to: US-09-132-619-7 from: 1 to: 2330

14 PheProLeuThrPhelySerSeraLaserProArgLys...PheLeuTh 29  
||||||| |||||:|||||: ||| ||| :|||  
868 TTTCCTGCAATTCAAACATCCATCTGCTCCAAAGGCCTACCTCTG 917

29 rAsnValThrGlyCysPheAlaThrValThrArgileProLeu 44  
:||||| :||| :||| :|||||:|||||  
918 AACCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 963

seq\_name: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:US-08-470-179-139

## seq\_documentation\_block:

; Sequence 139, Application US/08470179  
; Patent No. 5645994  
; GENERAL INFORMATION:  
; APPLICANT: Huang Ph.D, Wai Mun  
; TITLE OF INVENTION: Method and Compositions for  
; IDENTIFICATION OF SPECIES IN A SAMPLE  
; NUMBER OF SEQUENCES: 207  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Trask, Britt and Rossa  
; STREET: P.O. Box 2550  
; CITY: Salt Lake City  
; STATE: Utah  
; COUNTRY: USA  
; ZIP: 84110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/470,179  
; FILING DATE:

```
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sweigert Ph.D, Susan E.
; REGISTRATION NUMBER: 36,289
; REFERENCE/DOCKET NUMBER: 2601
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 801-532-1922
; TELEFAX: 801-531-9168
; INFORMATION FOR SEQ ID NO: 139:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 423 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: not relevant
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Leptospira borgpetersenii
; US-08-470-179-139

alignment_scores:
  Quality: 54.50      Length: 64
  Ratio: 1.652       Gaps: 1
  Percent Similarity: 51.562   Percent Identity: 21.875

alignment_block:
US-09-209-961-7 x US-08-470-179-139 ..
Align seg 1/1 to: US-08-470-179-139 from: 1 to: 423

11 PheCysIlePhePro..... 15
164 TTCAGGATTTTCCTCCGCGGTCCTCATGATGAGCGCAAGAAATTC 213
16 .....LeuThrPheLysSerSerAlaSerProArgLysP 27
214 GGTCCTATAGCGCGGATACCCCTGGCGCTTATCGGTACACCGAGCAAG 263
27 heLeuThrAsnValThrGlyCysCysPheAlaThrValThrArgIlePro 43
264 ACTTGAAAAAGTCGCGGAAGAATTGTTACGGGACATTGAAAAAGAGACCG 313
44 LeuSerAsnLysValLeuThrAlaValAspArgSerLeuArg 57
314 TTAGTTCTCGCCTAATACGATGATACAAACAAACACCGG 355

seq_name: /cgn2_6/ptodata/1/ina/3C_COMB.seq:US-08-125-462-8

seq_documentation_block:
; Sequence 8, Application US/08125462
; Patent No. 5840840
; GENERAL INFORMATION:
; APPLICANT: Rybak, Susanna M.
; APPLICANT: Youle, Richard J.
; APPLICANT: Newton, Dianne L.
; APPLICANT: Nicholls, Peter J.
; TITLE OF INVENTION: Selective RNase Cytotoxic Reagents
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Khourie and Crew
; STREET: Steuart Street Tower, One Market Plaza
; CITY: San Francisco
; STATE: California
; COUNTRY: US
; ZIP: 94105-1493
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/125,462
; FILING DATE: 22-SEP-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Ellen L.
; REGISTRATION NUMBER: 32,762
; REFERENCE/DOCKET NUMBER: 15280-110-3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 543-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 381 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1..381
; OTHER INFORMATION: /standard_name= "Human RNase Gene"
; US-08-125-462-8

alignment_scores:
  Quality: 54.00      Length: 80
  Ratio: 1.317       Gaps: 4
  Percent Similarity: 51.250   Percent Identity: 28.750

alignment_block:
US-09-209-961-7 x US-08-125-462-8 ..
Align seg 1/1 to: US-08-125-462-8 from: 1 to: 381

2 LysCysThrLeuValPheGlnSerArgPheCysIlePheProLeuThrPh 18
125 CGGTACACATTCGTTTCATGACCGCTGG.....TTGACGTT 162
18 eLysSerSerAlaSerProArgLysPheLeu..... 28
163 CAGAACGTTTCCTCCAGGAAAAAGTTACTTCGAAACACGTCAGGGTAA 212
29 .....ThrAsnValThrGlyCysCysPheAla..... 37
213 CTGCTACAAATCACTTCTTATGATATCATCATGACTGCCGCTGTGACTA 262
38 .....ThrValThrArgIleProLeuSer..... 45
263 ACGGTTCTCGTTACCGGAACGCGCTTACCGTACTTCTCGAAGAACGCT 312
46 ....AsnLysValLeuThrAlaValAspArgSerLeuArg 57
313 CATATCATCGTTGTTGCGAAGGTTCTCCGTACGTTCCGG 352

seq_name: /cgn2_6/ptodata/1/ina/3D_COMB.seq:US-08-891-848-8

seq_documentation_block:
; Sequence 8, Application US/08891848
; Patent No. 5955073
; GENERAL INFORMATION:
; APPLICANT: Rybak, Susanna M.
; APPLICANT: Youle, Richard J.
; APPLICANT: Newton, Dianne L.
; APPLICANT: Nicholls, Peter J.
; TITLE OF INVENTION: Selective RNase Cytotoxic Reagents
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
```



```
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: ARCD:155/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (713) 789-2679
; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1094 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-431-080-17

alignment_scores:
  Quality: 58.00      Length: 36
  Ratio: 2.417       Gaps: 2
Percent Similarity: 66.667 Percent Identity: 44.444

alignment_block:
US-09-209-961-7 x US-08-431-080-17 ..
Align seg 1/1 to: US-08-431-080-17 from: 1 to: 1094

7 PheGlnSerArgPheCys.....IlePheProLeuThrPheLysSerSe 21
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
70 TTCTCCTCGATTGTTGTTATCATCTTCCACAACTTTCCTGGGCTC 119

21 rAlaSerProArgLysPheLeuThrAsnValThrGlyCysCysPheAla 38
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
120 CACT.....TTTTGATTGGTGTGCTTGAATTGTTATCTGA 157

38 hrValThr 40
|||||:
158 CTCCTCACT 165

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-431-080-30

seq_documentation_block:
; Sequence 30, Application US/08431080
; Patent No. 5698686
; GENERAL INFORMATION:
; APPLICANT: Gottschling, Daniel E.
; APPLICANT: Singer, Miriam S.
; TITLE OF INVENTION: Telomerase Compositions and Methods
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TEXAS
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/431,080
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: SN 08/326,781
; FILING DATE: October 20, 1994
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
```

```
; REFERENCE/DOCKET NUMBER: ARCD:155/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (713) 789-2679
; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1094 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-431-080-30

alignment_scores:
  Quality: 58.00      Length: 36
  Ratio: 2.417       Gaps: 2
Percent Similarity: 66.667 Percent Identity: 44.444

alignment_block:
US-09-209-961-7 x US-08-431-080-30/rev ..
Align seg 1/1 to reverse of: US-08-431-080-30 from: 1 to: 1094

7 PheGlnSerArgPheCys.....IlePheProLeuThrPheLysSerSe 21
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1025 TTCTCCTCGATTGTTGTTATCATCTTCCACAACTTTCCTGGGCTC 976

21 rAlaSerProArgLysPheLeuThrAsnValThrGlyCysCysPheAla 38
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
975 CACT.....TTTTGATTGGTGTGCTTGAATTGTTATCTGA 938

38 hrValThr 40
|||||:
937 CTCCTCACT 930

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-938-534-17

seq_documentation_block:
; Sequence 17, Application US/08938534
; Patent No. 5916752
; GENERAL INFORMATION:
; APPLICANT: Gottschling, Daniel E.
; APPLICANT: Singer, Miriam S.
; TITLE OF INVENTION: Telomerase Compositions and Methods
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TEXAS
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/938,534
; FILING DATE: 26-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/431,080
; FILING DATE:
; APPLICATION NUMBER: SN 08/326,781
; FILING DATE: October 20, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: ARCD:155/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
```

```
; TELEFAX: (713) 789-2679
; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1094 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-938-534-17

alignment_scores:
  Quality: 58.00      Length: 36
  Ratio: 2.417       Gaps: 2
  Percent Similarity: 66.667   Percent Identity: 44.444

alignment_block:
  US-09-209-961-7 x US-08-938-534-17 ..
  Align seg 1/1 to: US-08-938-534-17 from: 1 to: 1094

7 PheGlnSerArgPheCys.....IlePheProLeuThrPheLysSerSe 21
  |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
70 TTCCTCTCGATTGTGTATCATCTTCCACAACTTTCTCGGGCTC 119

21 rAlaSerProArgLysPheLeuThrAsnValThrGlyCysPheAla 38
  |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
120 CACT.....TTTTGATTGGTGTGCTTGAATTGTATCTGA 157

38 hrValThr 40
  |||:|||||
158 CTCCTCACT 165

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-938-534-30

seq_documentation_block:
; Sequence 30, Application US/08938534
; Patent No. 5916752
; GENERAL INFORMATION:
; APPLICANT: Gottschling, Daniel E.
; APPLICANT: Singer, Miriam S.
; TITLE OF INVENTION: Telomerase Compositions and Methods
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TEXAS
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/938,534
; FILING DATE: 26-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; FILING DATE: 08/431,080
; APPLICATION NUMBER: SN 08/326,781
; FILING DATE: October 20, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: ARCD:155/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (713) 789-2679
; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 30:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1094 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-938-534-30

alignment_scores:
  Quality: 58.00      Length: 36
  Ratio: 2.417       Gaps: 2
  Percent Similarity: 66.667   Percent Identity: 44.444

alignment_block:
  US-09-209-961-7 x US-08-938-534-30 rev ..
  Align seg 1/1 to reverse of: US-08-938-534-30 from: 1 to: 1094

7 PheGlnSerArgPheCys.....IlePheProLeuThrPheLysSerSe 21
  |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
1025 TTCCTCTCGATTGTGTATCATCTTCCACAACTTTCTCGGGCTC 976

21 rAlaSerProArgLysPheLeuThrAsnValThrGlyCysPheAla 38
  |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
975 CACT.....TTTTGATTGGTGTGCTTGAATTGTATCTGA 938

38 hrValThr 40
  |||:|||||
937 CTCCTCACT 930

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-09-132-619-5

seq_documentation_block:
; Sequence 5, Application US/09132619B
; Patent No. 5958697
; GENERAL INFORMATION:
; APPLICANT: Shan, Bei
; APPLICANT: Nitta, Masahiro
; TITLE OF INVENTION: CYP7 Promoter-Binding Factors
; FILE REFERENCE: T97-013
; CURRENT APPLICATION NUMBER: US/09/132,619B
; CURRENT FILING DATE: 1998-08-11
; EARLIER APPLICATION NUMBER: 60/067,708
; EARLIER FILING DATE: 1997-12-08
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 3251
; TYPE: DNA
; ORGANISM: human
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (208)..(1830)
; US-09-132-619-5

alignment_scores:
  Quality: 62.50      Length: 32
  Ratio: 2.717       Gaps: 1
  Percent Similarity: 71.875   Percent Identity: 43.750

alignment_block:
  US-09-209-961-7 x US-09-132-619-5 ..
  Align seg 1/1 to: US-09-132-619-5 from: 1 to: 3251

14 PheProLeuThrPheLysSerSerAlaSerProArgLys...PheLeu 29
  |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
833 TTCCTCTGCAATTCAAACATCCACTCTGCCTCCAAAGGCGCTACTCTG 882

29 rAsnValThrGlyCysCysPheAlaThrValThrArgIleProLeu 44
  |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
883 AACCATGCTGCTTGCCTCTACACACATATGACAGAAGTCCCTTTG 928
```

seq\_name: /cgn2\_6/ptodata/1/ina/5C\_COMB.seq:US-08-589-711-1

seq\_documentation\_block:  
; Sequence 1, Application US/08589711  
; Patent No. 5853987  
; GENERAL INFORMATION:  
; APPLICANT: Guo, Magnus  
; APPLICANT: Hook, Magnus  
; TITLE OF INVENTION: DECORIN BINDING PROTEIN COMPOSITIONS  
; TITLE OF INVENTION: AND METHODS OF USE  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: P.O. Box 4433  
; CITY: Houston  
; STATE: Texas  
; COUNTRY: United States of America  
; ZIP: 77210  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/589,711  
; FILING DATE: Concurrently Herewith  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US SN 08/427,023  
; FILING DATE: 24-APR-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kitchell, Barbara S.  
; REGISTRATION NUMBER: 33,928  
; REFERENCE/DOCKET NUMBER: TAMK:158  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (512) 418-3000  
; TELEFAX: (512) 474-7577  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2653 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1471..2031  
US-08-589-711-1

alignment\_scores:  
Quality: 61.50 Length: 57  
Ratio: 1.984 Gaps: 2  
Percent Similarity: 54.386 Percent Identity: 31.579  
alignment\_block:  
US-09-209-961-7 x US-08-589-711-1 ..  
Align seg 1/1 to: US-08-589-711-1 from: 1 to: 2653  
1 MetLysCysThrLeuValPheGlnSerArg..... 10  
2244 ATAAATGTAATTATTATTTCACAAATAATTAATATCCTTTTGA 2293  
11 .....PheCysilePheProLeuThrPheLysSerSerAlaSerProA 25  
2294 TGTATTGGGAATTTCTTCCTTAGACTTTAAATCAAGACTGTCGTA 2343  
25 rgLysPheLeuThrAsnValThrGlyCysCysPheAlaThrValThrArg 41  
2344 GCACCTATTATTATCATCAACAAATGC.....ACAAACCCGA 2387  
42 ilPLeuSerAsnLysVal 48

... |||.....  
2388 CTTTACCTTAACCTCTGTATT 2408

seq\_name: /cgn2\_6/ptodata/1/ina/5C\_COMB.seq:US-08-739-485-6

seq\_documentation\_block:  
; Sequence 6, Application US/08739485  
; Patent No. 5863898  
; GENERAL INFORMATION:  
; APPLICANT: Goli, Surya K.  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Bandman, Olga  
; TITLE OF INVENTION: NOVEL HUMAN LIM PROTEINS  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: US  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/739,485  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0142 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1336 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: Consensus  
; CLONE: Consensus  
US-08-739-485-6

alignment\_scores:  
Quality: 58.50 Length: 39  
Ratio: 2.659 Gaps: 1  
Percent Similarity: 56.410 Percent Identity: 35.897  
alignment\_block:  
US-09-209-961-7 x US-08-739-485-6/rev ..  
Align seg 1/1 to reverse of: US-08-739-485-6 from: 1 to: 1336  
21 SerAlaSerProArgLysPheLeuThrAsnValThrGlyCysPheAl 37  
334 AGTTTCTCCCTCAATTTTGACACACAGCTGGTGAGCTGCTGCTTAA 285  
37 aThrValThrArgIleProLeuSerAsnLysValLeuThrAlaValAsp 54  
284 TCCTGTCTCTGGCATCAGCATGAGTCATGGACTC..... 251  
54 rgSerLeuArgCysPro 59  
|||||

```
alignment_scores:
```



```

alignment_scores:      Quality: 53.50      Length: 67
                       Ratio: 1.672      Gaps: 3
                       Percent Similarity: 47.761      Percent Identity: 29.851

alignment_block:
US-09-209-961-7 x US-08-009-261-4 ..

Align seg 1/1 to: US-08-009-261-4 from: 1 to: 492

1 MetLysCysThrLeuValPheGlnSerArgPheCys.....IlePh 14
   :::::  ::::|  |||||:::|  ::::|  ::::|  ::::|  ::::|
90 CTACGAGTGTCTTCATCATCATCTAGGTATCTGACATTAGGATTCTT 139
   ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
14 eProLeuThrPhe..... 18
   ||||  :::::

140 CCTCAGAATTATTCTCGAATAATCCGACTCCCTTCACTGTGCATACCAAT 189
   ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
19 .....LysSerSerAlaSerProArgLysPheLeuThrAsn 30
   :::::  :::::  ::::|  ||||  |||||  |||||  |||||  ::::
190 TGTGAGAATCTCTCTGGAATCATATATCTACCGCAAATTCCTGTAAACAT 239
   ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
31 ValThrGlyCysCysPheAlaThrValThrArgIleProLeuSerAsnly 47
   |||||  ||||  |||||  ::::|  ::::|  ::::|  ::::|  ::::|
240 ...ACAGAGATTGTCTCTTCATGCACACAGCACTTCTCTTATGAATGA 286
   ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|

47 s 47
:
287 g 287

```

```

seq_documentation_block:
: Sequence 4, Application US/08009261
: Patent No. 5350672
: GENERAL INFORMATION:
: APPLICANT: Oberst, Richard D.
: APPLICANT: Gwaltney, Sharon M.
: APPLICANT: Hays, Michael P.
: TITLE OF INVENTION: Specific DNA Primers and Method to Use
: TITLE OF INVENTION: Same to Detect Eperythrozoon suis
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Tilton, Fallon, Lungmus and Chestnut
: STREET: 100 S. Wacker Drive, Suite 960
: CITY: Chicago
: STATE: Illinois
: COUNTRY: USA
: ZIP: 60606-4002
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/009,261
: FILING DATE: 19930122
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: Fentress, Susan B.
: REGISTRATION NUMBER: 31,327
: REFERENCE/DOCKET NUMBER: 92061A
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (312)456-8000
: TELEFAX: (312)456-7776
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 492 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: unknown
: TOPOLOGY: unknown
: MOLECULE TYPE: DNA (genomic)
: PS-08-009-261-4

```



Strid	Orig	ZScore	E_Score	Len	Documentation
gb_gss7:AQ500736	+ 76.00	166.50	1.49	543	AQ500736 V53G6 mTn-3xHA/lacZ In
gb_gss7:AQ503361	- 71.00	156.50	1.49	561	AQ503361 V57C10 mTn-3xHA/lacZ In
gb_gss7:AQ502889	+ 70.00	153.00	7.05	505	AQ502889 V48G6 mTn-3xHA/lacZ In
gb_gss7:AQ491934	+ 71.00	150.83	11.10	816	AQ491934 M271D4 mTn-3xHA/lacZ In
gb_gss7:AQ491969	+ 71.00	150.59	11.45	837	AQ491969 M614E9 mTn-3xHA/lacZ In
gb_est8:C07961	- 66.50	149.64	12.94	300	C07961 C07961 Yuj1 Kohara unpub
gb_est16:C36594	- 66.50	149.64	12.94	300	C36594 C36594 Yuj1 Kohara unpub
gb_est16:C57408	- 66.50	149.64	12.94	300	C57408 C57408 Yuj1 Kohara unpub
gb_est36:AV178364	- 66.50	149.64	12.94	300	AV178364 AV178364 Yuj1 Kohara unpub
gb_est36:AV182172	- 66.50	149.64	12.94	300	AV182172 AV182172 Yuj1 Kohara unpub
gb_est36:AV192309	- 66.50	149.64	12.94	300	AV192309 AV192309 Yuj1 Kohara unpub
gb_est36:AV184335	- 66.50	149.64	12.94	300	AV184335 AV184335 Yuj1 Kohara unpub
gb_est36:AV185677	- 66.50	149.64	12.94	300	AV185677 AV185677 Yuj1 Kohara unpub
gb_est5:D72647	- 66.50	148.94	14.15	323	D72647 CELK106CVR Yuj1 Kohara unpub
gb_est17:AAQ605665	+ 69.00	147.94	16.10	372	AAQ605665 fa2ze08.sl Exkerpest s
gb_est12:Rf6521	+ 66.00	147.94	16.40	322	Rf6521 yd87d06.r1 Soares fetal
gb_est36:AI897437	- 68.00	147.35	17.36	557	AI897437 ESTr266880 tomato ovary
gb_est11:AA248089	- 65.00	147.09	17.94	270	AA248089 cl0567.seq.F Human fet
gb_gss7:AQ501875	+ 66.50	145.38	22.33	471	AQ501875 V13c4 mTn-3xHA/lacZ In
gb_est16:C39191	- 64.50	144.91	23.72	300	C39191 C39191 Yuj1 Kohara unpub
gb_gss3:AO210699	+ 66.00	144.40	25.32	461	AO210699 HS-2226_BI_H05_MF Cl
gb_est38:AW037261	+ 66.50	144.27	25.76	530	AW037261 ca03a04.yl C elegans f
gb_gss3:AO202723	+ 61.00	143.97	26.77	138	AO202723 l(2)k03402 Drosophila
gb_est38:AW049581	- 66.00	143.89	27.06	487	AW049581 UI-M-BH1-and-a-04-O-U
gb_est28:IA487234	- 66.00	143.81	27.33	491	IA487234 ESTr445556 tomato ovary
gb_est29:AI587703	+ 66.00	143.36	28.96	515	AI587703 vt17a05.yl Barstead md
gb_est22:AI800058	- 64.00	142.89	30.75	328	AI800058 ui-R-CJ-kz-f-10-O-U.s.
gb_est22:AI006582	+ 67.00	142.48	32.41	726	AI006582 ue15f01.yl Sugano mous
gb_est21:AA939988	+ 64.00	141.99	34.54	362	AA939988 vq65f12.sl Knowles Sol
gb_est22:AI004472	+ 64.00	141.70	35.81	372	AI004472 UI-R-CJ-kb-d-03-O-U.s.
gb_est25:AA137649	+ 65.00	141.62	36.19	482	AA137649 u21g11.yl Sugano mous
gb_est14:AA397624	+ 64.50	141.56	36.47	428	AA397624 zt79d05.r1 Soares test
gb_est28:AI516812	+ 65.00	141.48	36.83	489	AI516812 LD42885.Sprime LD bro
gb_est27:AI429316	- 64.00	141.43	37.10	383	AI429316 mm4e04.xl Stratagene
gb_est36:AV175812	- 63.00	141.37	37.38	300	AV175812 AV175812 Yuj1 Kohara u
gb_gss6:B75805	+ 64.50	141.28	37.82	441	B75805 RPC11-l12H13.TU RPC1-1
gb_gss9:AO618111	+ 65.00	141.00	39.21	515	AO618111 HS_5168_BI_C07_A7A_RP
gb_gss4:AO118407	+ 65.00	140.72	40.60	530	AO118407 HS_3007_A2_D03_T7 CIT
gb_est22:AI044868	+ 64.00	140.31	42.80	431	AI044868 HOMS0Y328 Human brain c
gb_est22:AI045037	- 64.00	140.25	43.16	434	AI045037 UI-R-CJ-kf-d-06-O-U.s.
gb_gss8:ANCS0397	+ 66.00	140.22	43.30	718	AQ513397 RPC11-l133cl17.TU RPC1-
gb_gss1:CN500136	+ 67.50	139.73	46.11	1101	AL072874 Drosophila melanogast

4

14 PheProLeuThrPhe.....LySerSerAlaSerProArgLy 26  
||||| |||  
58 TTCGCCCTCGCTTCGTTTCGAAACATACAGATGGGSCAGACCCACGTCG 107  
  
26 sPheLeuThrAsnValThrGlyCysCysPheAlaThrValThrArgIleP 43  
:|||:||||| ||| |||:  
108 A.....AACACGTCGCGACTTCGCTCTTCGCGACTTTTACCAAGTTCA 151  
  
43 toLeuSerAsnLySvalLeuThrAlaValAspArg 54  
|||:||||| |||:  
152 TCCTTACTCGAGGGGTCTTGTAAACACAGATCGA 186

seq\_name: gb\_gss7:AQ491969

seq\_documentation\_block:

LOCUS AQ491969 837 bp DNA GSS 23-APR-1999

DEFINITION M61459 mfn-3xHA/lacZ Insertion Library Saccharomyces cerevisiae genomic 5', genomic survey sequence.

ACCESSION AQ491969

NID 94677843

VERSION AQ491969.1 GI:4677843

KEYWORDS GSS.

SOURCE baker's yeast.

ORGANISM Saccharomyces cerevisiae

Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.

REFERENCE 1 (bases 1 to 837)

AUTHORS Ross-Macdonald,P., Roemer,T., Coelho,P.S.R., Agarwal,S., Kumar,A., desEtages,S.A., Cheung,K.-H., Sheehan,A., Symoniatidis,D., Jansen,R., Umanaky,L., Heidtman,M., Nelson,K., Iwasaki,H., Kanada,D., Lugo,R., Hager,K., Miller,P., Roeder,G.S. and Snyder,M.

TITLE Large-Scale Analysis of the Yeast Genome by Transposon Tagging and

Gene Disruption  
Unpublished (1999)

Contact: Kumar A  
Michael Snyder, Dept. of Mol. Cell. and Dev. Biology  
Yale University  
P.O. Box 208103, New Haven, CT 06520-8103, USA  
Tel: 203 432 9949  
Fax: 203 432 6161  
Email: anuj.kumar@yale.edu

Sequence corresponds to the extreme 5' end of mTn-3xHA/lacZ (ending in TCAGACCCC) followed by native *Saccharomyces cerevisiae* genomic sequence immediately upstream; generated by single read using sequencing primer indicated above; enables determination of site of mTn-3xHA/lacZ insertion.

```

FEATURES             source
    source
        1. .837
            /organism="Saccharomyces cerevisiae"
            /db_xref="taxon:4932"
            /clone.lib="mtfN-3xHA/lacZ Insertion Library"
            /lab_host="E. coli"
            /note="vector: pHS56-Sal; A yeast genomic DNA library
                    (lacking mitochondrial DNA) was prepared in pHS56-Sal;
                    genomic DNA was size-fractionated (DNA of roughly 2-3 kb
                    in length) prior to cloning. This library was
                    subsequently mutagenized with a mTfN-3xHA/lacZ
                    minitransposon containing lacZ, URA3, and tet resistance."
    Seq primer: GGCGTCTCTTCTCTTTGGGAGTAC
    Class: transposon-tagged.
    Location/Qualifiers

```

```

ORIGIN
alignment_scores:
  Quality: 71.00
  Ratio: 2.290
  Percent Similarity: 60.784
  Length: 51
  Gaps: 2
  Percent Identity: 39.216
alignment_block:
  US-09-209-961-7 x AQ491969 ..

```

Align seg 1/1 to: A0491969 from: 1 to: 837

```

9 SerArgPheCysIlePheProLeuThrPhe.....LysSerSe 21
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
102 AGTCCTTTAGTGAATCCCTCGCTTCGTTCCGAACCATAGATGG 151
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
21 rAlaSerProArgLysPheLeuThrAsnValThrGlyCysCysPheAla 38
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
152 GGCAGACCCAGCTCGA.....AACACGTCGGCAGCTGCTGTTCCGCA 195
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
38 hrValThrArgIleProLeuSerAsnLysValLeuThrAlaValAspArg 54
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
196 CTTTACCAAGTTCATCTTACTGGAGGGTCTTGAAACACAGATCAG 245
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
55 Ser 55
   |||
246 ACG 248
   |||

```

seq\_name: gb\_est8:C07961

```

seq_documentation_block: 300 bp mRNA EST 12-SEP-1996
LOCUS C07961 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite embryo
DEFINITION Caenorhabditis elegans cDNA clone yk172ell 3', mRNA sequence.
ACCESSION C07961
NID 91533032
VERSION C07961.1 GI:1533032
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 300)
AUTHORS Kohara,Y., Motoshashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
Sano,M., Miyata,A. and Nishigaki,A.
TITLE Expression map of the C.elegans genome
JOURNAL Unpublished (1996)
COMMENT On Nov 29, 1993 this sequence version replaced gi:637700.

```

```

FEATURES
    source
        1..300
            /organism="Caenorhabditis elegans"
            /strain="N2"
            /db_xref="taxon:6239"
            /clone="yk172ell"
            /clone_lib="Yuji Kohara unpublished cDNA:Strain N2
            hermaphrodite embryo"
            /sex="hermaphrodite"
            /dev_stage="embryo"
            Location/Qualifiers
                75 a 70 c 77 g 78 t
BASE COUNT 75 a 70 c 77 g 78 t
ORIGIN

```

alignment\_scores:  
 Quality: 66.50 Length: 42  
 Ratio: 2.293 Gaps: 1  
 Percent Similarity: 69.048 Percent Identity: 35.714

alignment\_block:

US-09-209-961-7 x C07961/rev ..

Align seg 1/1 to reverse of: C07961 from: 1 to: 300

```

1 MetLysCysThrLeuValPheGlnSerArgPhe...CysIlePheProLe 16
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
194 ATGGCTTCAACTCTGAGTATCCAGTGGAGAACTGATCGGTGATGCCAAG 145
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

16 utrPhelySerSerAlaSerProArgLysPheLeuThrAsnValThrG 33
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
144 ATCTATCAGATCTACGAGGGAACCTCGCAATTCAGCGCATGTCATCTC 95
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

33 lyCysCysPheAlaThrValThrArg 41
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
94 GCGCATGCTTCTCGGACATTTCCGGC 69
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

seq\_name: gb\_est16:C36594

```

seq_documentation_block: 300 bp mRNA EST 10-SEP-1997
LOCUS C36594 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite embryo
DEFINITION Caenorhabditis elegans cDNA clone yk468d4 3', mRNA sequence.
ACCESSION C36594
NID 92372735
VERSION C36594.1 GI:2372735
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 300)
AUTHORS Kohara,Y., Motoshashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
Sano,M., Miyata,A. and Nishigaki,A.
TITLE Expression map of the C.elegans genome
JOURNAL Unpublished (1996)
COMMENT On Dec 30, 1996 this sequence version replaced gi:1529573.

```

```

Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers
    1..300
        /organism="Caenorhabditis elegans"
        /strain="N2"
        /db_xref="taxon:6239"
        /clone="yk468d4"
        /clone_lib="Yuji Kohara unpublished cDNA:Strain N2
        hermaphrodite embryo"
        /sex="hermaphrodite"
        /dev_stage="embryo"
        Location/Qualifiers
            75 a 69 c 78 g 78 t
BASE COUNT 75 a 69 c 78 g 78 t
ORIGIN

```

alignment\_scores:  
 Quality: 66.50 Length: 42  
 Ratio: 2.293 Gaps: 1  
 Percent Similarity: 69.048 Percent Identity: 35.714

alignment\_block:

US-09-209-961-7 x C36594/rev ..

Align seg 1/1 to reverse of: C36594 from: 1 to: 300

```

1 MetLysCysThrLeuValPheGlnSerArgPhe...CysIlePheProLe 16
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
195 ATGGCTTCAACTCTGAGTATCCAGTGGAGAACTGATCGGTGATGCCAAG 146
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

16 utrPhelySerSerAlaSerProArgLysPheLeuThrAsnValThrG 33
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
145 ATCTATCAGATCTACGAGGGAACCTCGCAATTCAGCGCATGTCATCTC 96
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

33 lyCysCysPheAlaThrValThrArg 41
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
95 GCGCATGCTTCTCGGACATTTCCGGC 70
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```







```

FEATURES
Source
Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers
1. .300
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/clone_lib="Yuji Kohara unpublished cDNA:Strain N2
hermaphrodite embryo"
/dev_stage="embryo"
74 a 69 c 80 g 77 t
BASE COUNT
ORIGIN

alignment_scores:
Quality: 66.50 Length: 42
Ratio: 2.293 Gaps: 1
Percent Similarity: 69.048 Percent Identity: 35.714

alignment_block:
US-09-209-961-7 x AV184335/rev ..
Align seg 1/1 to reverse of: AV184335 from: 1 to: 300
1 MetLysCysThrLeuValPheGlnSerArgPhe...CysIlePheProle 16
||||| ||||| :||| ||| |||||:||||
202 ATGGCTTCAACTCTGAGTATCAGTGGGAGAACTGCGTGCATGCCAAG 153
16 uThrPheLysSerAlaSerProArgLysPheLeuThrAsnValThrG 33
:|||||: :||| ||||| ||| :|||
152 ATCTATCAGATCTACGAGGGAACCTCGCAATTCAGCGCATGTCATCTC 103
33 lyCysCysPheAlaThrValThrArg 41
:|||||: :||| ||||| ||| :|||
102 CGCGATGCTCTCGGACATTCGCGC 77
seq_name: gb_est36:AV185677

seq_documentation_block:
LOCUS AV185677 300 bp mRNA EST 21-JUL-1999
DEFINITION AV185677 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite
embryo Caenorhabditis elegans cDNA clone yk681h6 3', mRNA sequence.
ACCESSION AV185677
NID 95565578
VERSION AV185677.1 GI:5565578
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
AUTHORS Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Mitsuki,H.,
Nishigaki,A., Motohashi,T., Zeng,Q., Watanabe,H., Sugimoto,A.,
Sano,M., Miyata,A., Mitani,Y., Iida,K., Uesugi,H., Sugiyama,Y. and
Nomoto,H.
TITLE Expressed genes in C.elegans
JOURNAL Unpublished (1999)
COMMENT On May 9, 1996 this sequence version replaced gi:1133045.

Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers
1. .300
/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/map="between D11S1765 and UGB"

FEATURES
source

```

```

/clone="yk681h6"
/clone_lib="Yuji Kohara unpublished cDNA:Strain N2
hermaphrodite embryo"
/sex="hermaphrodite"
/dev_stage="embryo"
75 a 68 c 77 g 75 t 5 others
BASE COUNT
ORIGIN

alignment_scores:
Quality: 66.50 Length: 42
Ratio: 2.293 Gaps: 1
Percent Similarity: 69.048 Percent Identity: 35.714

alignment_block:
US-09-209-961-7 x AV185677/rev ..
Align seg 1/1 to reverse of: AV185677 from: 1 to: 300
1 MetLysCysThrLeuValPheGlnSerArgPhe...CysIlePheProle 16
||||| ||||| :||| ||| |||||:||||
204 ATGGCTTCAACTCTGAGTATCAGTGGGAGAACTGATGCGTGCATGCCAAG 155
16 uThrPheLysSerAlaSerProArgLysPheLeuThrAsnValThrG 33
:|||||: :||| ||||| ||| :|||
154 ATCTATCAGATCTACGAGGGAACCTCGCAATTCAGCGCATGTCATCTC 105
33 lyCysCysPheAlaThrValThrArg 41
:|||||: :||| ||||| ||| :|||
104 GCGCATGCTCTCGGACATTCGCGC 79
seq_name: gb_est5:D72647

seq_documentation_block:
LOCUS D72647 323 bp mRNA EST 13-DEC-1995
DEFINITION CELK106CYR Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite
embryo Caenorhabditis elegans cDNA clone yk106c11 3', mRNA
sequence.
ACCESSION D72647
NID g1112362
VERSION D72647.1 GI:1112362
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
AUTHORS Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and
Tabara,H.
TITLE Toward an expression map of the C.elegans genome
JOURNAL Unpublished (1994)
COMMENT On May 8, 1995 this sequence version replaced gi:801481.

Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp
Insert length: 1990 Std Error: 0.00
High quality sequence stop: 307.
Location/Qualifiers
1. .323
/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/map="4"
/clone="yk106c11"
/clone_lib="Yuji Kohara unpublished cDNA:Strain N2
hermaphrodite embryo"
/sex="hermaphrodite"
/dev_stage="embryo"

```



---



OM of: US-09-209-961-9 to: GenEmbl.\* out\_format : pfs

Date: Dec 27, 1999 2:13 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-1998 CompuGen Ltd.

#### Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp  
-Q=Cgcn2.1YUSPTO.spool/US09209961/runat\_22121999\_101043\_17910/app\_query.fasta.1  
-DB=GenEmbl -QPMT=fastap -SUFFIX=rgs -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPEXT=0.000 -GAPOP=4.500  
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blossum62 -TRANS=human4.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=score -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pfs -NORM=stat -USER=US09209961 -NCPU=6 -ICPU=3 -NO\_XLPHY  
-WAIT -THREADS=1

#### Search information block:

Query: US-09-209-961-9  
Query length: 104  
Database: GenEmbl.\*  
Database sequences: 780561  
Database length: 2137953050  
Search time (sec): 1994.760000

#### score\_list:

Sequence	Strd Orig	zscore	EScore	Len	Documentation
gb_v1:AF085695	- 572.00	1150.26	4.5e-56	1768	AF085695 Porcine circovirus st
gb_v1:AF086834	- 572.00	1150.26	4.5e-56	1768	AF086834 Porcine circovirus st
gb_v1:AF086835	- 572.00	1150.26	4.5e-56	1768	AF086835 Porcine circovirus st
gb_v1:AF086836	- 572.00	1150.26	4.5e-56	1768	AF086836 Porcine circovirus st
gb_v1:AF118095	- 541.00	1096.59	4.4e-53	597	AF118095 Porcine circovirus ty
gb_v1:AF027217	- 541.00	1086.76	1.6e-52	1768	AF027217 Porcine circovirus st
gb_v1:AF055391	- 541.00	1086.76	1.6e-52	1768	AF055391 Porcine circovirus ty
gb_v1:AF055392	- 541.00	1086.76	1.6e-52	1768	AF055392 Porcine circovirus ty
gb_v1:AF109397	- 541.00	1086.76	1.6e-52	1768	AF109397 Bovine circovirus, c
gb_v1:AF147751	- 541.00	1086.76	1.6e-52	1768	AF147751 Porcine circovirus ty
gb_v1:AF109398	- 535.00	1074.47	7.6e-51	1768	AF109398 Porcine circovirus ty
gb_v1:PCAJ3185	- 532.00	1068.32	1.7e-51	1768	AF223185 Porcine circovirus ty
gb_v1:AF112862	- 531.00	1066.27	2.2e-51	1768	AF112862 Porcine circovirus ty
gb_v1:AF154679	- 522.00	1047.84	2.3e-50	1768	AF154679 Porcine circovirus, c
gb_v1:AF166528	- 522.00	1047.84	2.3e-50	1768	AF166528 Porcine circovirus c
gb_v1:AF055393	- 520.00	1043.74	3.9e-50	1767	AF055393 Porcine circovirus ty
gb_v1:AF055394	- 520.00	1043.74	3.9e-50	1767	AF055394 Porcine circovirus ty
gb_v1:AF117753	- 511.00	1025.30	4.1e-49	1768	AF117753 Porcine circovirus ty
gb_v1:AF109398	- 506.00	1015.06	1.5e-48	1768	AF109398 Porcine circovirus ty
gb_v1:AF071879	- 278.00	548.05	1.6e-22	1758	AF071879 Porcine circovirus, c
gb_v1:PCOMGEN	- 276.00	548.04	1.6e-22	1759	Y09921 Porcine circovirus comp
gb_v1:AF012107	- 276.00	543.94	2.7e-22	1759	AF012107 Porcine circovirus, c
gb_v1:PCU49186	- 276.00	543.94	2.7e-22	1759	U49186 Porcine circovirus, com
gb_v1:HS28C9R	- 74.00	145.60	4.14	319	Z58065 H.sapiens CpG island DNA
gb_v1:AF074399	+ 75.50	141.38	7.12	714	AF074399 Medicago truncatula in
gb_v1:HSU79241	+ 77.50	139.32	9.28	1410	U79241 Human clone 23759 mRNA,
gb_v1:AF055579	+ 71.50	138.48	10.33	398	AU025579 Rattus norvegicus, OTS
gb_v1:AF012107	+ 71.50	138.06	10.91	417	M61842 Human lactase phlorizin
gb_v1:AF103777	+ 72.50	137.12	12.31	580	AF103777 Pepper mild mottle vir
gb_v1:AF118935	+ 69.00	135.71	14.74	307	AF118935 Mus musculus AE0018 in
gb_v1:FE235274	- 71.50	135.02	16.10	583	AU235274 Itersonilia perplexans
gb_v1:AF053170	- 72.00	133.95	18.48	735	AF053170 Trifolium pallescens l
gb_v1:MTVCP	+ 69.50	132.80	21.41	474	D13367 Tobacco mosaic virus CP
gb_v1:HSAL32WE9	+ 67.00	130.58	28.46	344	Z24643 H. sapiens (D16S517) DNA
gb_v1:CMU45720	- 68.50	130.53	28.66	485	AU45720 Candida mesenterica 26S
gb_v1:GGA240695	- 68.50	130.29	29.56	499	AU240695 Gallus gallus DNA for
gb_v1:AF152942	- 68.50	129.74	31.70	530	AF152942 Clavospora opuntiae st
gb_v1:MTVCOATPRA	+ 69.50	129.56	32.45	678	L35073 Tobacco mosaic virus coa
gb_v1:MTVCOATPRO	+ 69.50	129.53	32.56	680	L35074 Tobacco mosaic virus coa
gb_v1:AB012214	+ 78.50	129.45	32.89	5253	AB012214 Rattus rattus mRNA fo
gb_v1:PMNVICP	+ 69.50	129.28	33.62	699	X72587 Pepper mild mottle virus
gb_v1:EL3985	+ 73.50	129.08	34.49	1766	EL3985 Arabidopsis thaliana mR

gb\_ov:MIEFCYTB + 71.50 128.93 35.19 1143 ! X74253 E.fastuosus mitochon  
gb\_pr1:D87058 + 70.00 128.75 35.99 830 ! D87058 Human DNA, endogenous  
gb\_ro:MMVH0001 + 66.00 128.69 36.26 338 ! X53338 Mouse mRNA for the V

seq\_name: gb\_v1:AF085695

#### seq\_documentation\_block:

LOCUS AF085695 1768 bp DNA circular VRL 30-SEP-1998  
DEFINITION Porcine circovirus strain 412, complete genome.

ACCESSION AF085695

NID 93688362

VERSION AF085695.1 GI:3668362

KEYWORDS porcine circovirus.

SOURCE Porcine circovirus

ORGANISM Viruses; ssDNA viruses; Circoviridae; Circovirus.

REFERENCE 1 (bases 1 to 1768)

AUTHORS Wang,L., Willson,P., Chow,B., Gibbons,E. and Babiuk,L.

TITLE Emergence of a new porcine circovirus

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1768)

AUTHORS Wang,L., Willson,P., Chow,B., Gibbons,E. and Babiuk,L.

TITLE Direct Submission

JOURNAL Submitted (21-AUG-1998) Veterinary Infectious Disease Organization,  
University of Saskatchewan, 120 Veterinary Road, Saskatoon,  
Saskatchewan S7N 5E3, Canada

#### FEATURES

source

1..1768 /organism="porcine circovirus"

/strain="412"

/db\_xref="taxon:46221"

51..995 /codon\_start=1

/product="P35.8"

/protein\_id="AAC61860.1"

/db\_xref="PID:g3668363"

/db\_xref="GI:3668363"

/translation="MPSKNGRSGPQPKRWVFLNPNSEDEKKIRELPISLFDYFI

VEEGNEERTPHLQGFANFKOTENKVWVUGARCHEKAGTKDOONKEIYSKREN

LIIEGAPRSQSQSRLSTAVSLGSLVTVAKQHPVTVFNFRGLAEILKVSQGM

QRDKNTNHFIVGPPCGSKSWANFANPETTYWPKPKNWDGYSGEVWVYDDFY

GWLPMDLLRLCDRYPLTVTKTGTVPLFLARSLITISNOTPLEWYSSTAVPAVALYR

RITSLVFNKATKQSGEEGQFVTLSPPCPEFFEYNY"

complement(357..671)

/codon\_start=1

/product="P12.1"

/protein\_id="AAC61865.1"

/db\_xref="PID:g3668368"

/db\_xref="GI:3668368"

/translation="MTIIPLVRFNFPVCGFRVCKISSPFAFTTPRPHNEVIGFPI

TLLHFFAHPKFSQPAEIDKRYRVLNCHGNHQPALQQGTHSSRQVTPLSLRSSSTF

NR"

complement(386..565)

/codon\_start=1

/product="P6.6"

/protein\_id="AAC61866.1"

/db\_xref="PID:g3668369"

/db\_xref="GI:3668369"

/translation="MKTCLVFQSRFCIFPLTFKSSASPRKFLNVTGCCFATVTRIP

SNKLVTAVDRLRCP"

553..732 /codon\_start=1

/product="P6.6"

/protein\_id="AAC61861.1"

/db\_xref="PID:g3668364"

/db\_xref="GI:3668364"

/translation="MTLSWGLGVVVGKANGLLILQTRKPTGHNHKTSGGMVTKVKKW

LILLMTFMACRGMV"

P33..970 /codon\_start=1

/product="P5.0"

/protein\_id="AAC61862.1"

/db\_xref="PID:g3668365"

/db\_xref="GI:3668365"  
/translation="MVLNCCPSRSLSESDYFLGLECYKTHGGRGVPVRHPFPPM  
P"

CDS

1016..1177

/codon\_start=1

/product="P6.2"

/protein\_id="AAC61863.1"

/db\_xref="PID:g3668366"

/db\_xref="GI:3668366"

/translation="MVFIIHGFKVGVEKIFSELYIHGYTDIWLVTTFVFSADA  
YVHISKGL"

CDS

complement(1034..1735)

/codon\_start=1

/product="P27.9"

/protein\_id="AAC61864.1"

/db\_xref="PID:g3668367"

/db\_xref="GI:3668367"

/translation="MTYPRRYRRRRHRRSHLQILRRPWLPHRRHYRWRKNGI  
FNTLSRTFGYTVKRTVTTPSWAVDMRKNIDFVPPGGTGNKISIPFYYIRKVK  
VEFWPCSPITQDGRGVGSTAVILDNFTKATALTYPVYNSRHTIPOPFSYHSRY  
FTPKPVLDSIDYFQPNKRNQLWRLQTSNDVHVLGTAFENSKYDQDYNIRVTMY  
VQREFNLKDPPLP"

BASE COUNT  
ORIGIN

463 a 362 c 481 g 462 t

alignment\_scores:

Quality: 572.00 Length: 104

Ratio: 5.500 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-209-961-9 x AF085695/rev ..

Align seg 1/1 to reverse of: AF085695 from: 1 to: 1768

1 MetValThrIleProLeuValPheArgTTPPheProValCysGlyPh 17

|||||

671 ATGGTAACCATCCACCACCTTGTGTTTAGTGGTTCCAGTATGTTT 622

|||||

17 eArgValCysLysIleSerSerProPheAlaPheThrProArgTTP 34

|||||

621 CCGGGTTTGCAAAATAGCAGCCATTTGCTTTTACCACACCCAGGTGC 572

|||||

34 roHisAsnGluValTyrIleGlyPheProIleThrLeuLeuHisPhePro 50

|||||

571 CCACAATGAAGTGTACATGTGTTTCCCAATCAGCTTTTGCATTTTCCC 522

|||||

51 AlaHisPheGlnLysPheSerGlnProAlaGluIlePheAspLysArgTy 67

|||||

521 GCTCACTTTCAAAAGTTCAGCAGCCCGCGGAAATTTTGACAAAGTTA 472

|||||

67 rArgValLeuLeuCysAsnGlyHisGlnAsnProAlaLeuGlnGlyT 84

|||||

471 CAGGTGCTGCTTTGCAAGGTCACAGATCCCGCTCTCCAACAGTTA 422

|||||

84 hrHisSerSerArgGlnValThrProLeuSerLeuArgSerArgSer 100

|||||

421 CTCACAGCAGTAGACAGGTGCTGCTGCTTTCAGATCGAGGAGCTCC 372

|||||

101 ThrPheAsnLys 104

371 ACATTCATAAG 360

seq\_name: gb\_v1:AF086834

seq\_documentation\_block:

LOCUS AF086834 1768 bp DNA circular VRL 29-SEP-1998

DEFINITION Porcine circovirus strain B9, complete genome.

ACCESSION AF086834

NID g3661515

VERSION AF086834.1 GI:3661515

KEYWORDS

SOURCE porcine circovirus.

ORGANISM porcine circovirus

REFERENCE 1 (bases 1 to 1768) Viruses; ssDNA viruses; Circoviridae; Circovirus.

AUTHORS Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.

JOURNAL Emergence of a new porcine circovirus

REFERENCE 2 (bases 1 to 1768) Unpublished

AUTHORS Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.

JOURNAL Direct Submission

TITLE Submitted (26-AUG-1998) VIDO, University of Saskatchewan, 120

JOURNAL Veterinary Road, Saskatoon, Saskatchewan S7N 5E3, Canada

FEATURES Location/Qualifiers

source

1..1768

/organism="porcine circovirus"

/strain="B9"

/db\_xref="taxon:46221"

51..995

CDS

/codon\_start=1

/product="P35.8"

/protein\_id="AAC61737.1"

/db\_xref="PID:g3661516"

/db\_xref="GI:3661516"

/translation="MPSKNGRSGPQPHKRVFTLNPNSEDERKKIRELPISLFDYFI  
VGEENEGRTPLHQGFANFVKQTFNKKWYLGARQHVTFVKNFGRGLAEKLVKSGKM  
LLIEGAPRSQGRSDLSSTAVSTLLESGILVTAQHPVTFVKNFGRGLAEKLVKSGKM  
OKRDWKTNVHIVGPPCGGSKWAANFANPETTYWKPKNWDGYSYHKEKVVYIDDFY  
GWLFDLLRLCDRIPLTVKTKGTFVPLKARSILITSNQTPLEWTSYSTAVFAVEALYR  
RITSLVFNKNVTEQSTEEGGQFVTLSPPCPEFFYEINY"

complement(1034..1735)

/codon\_start=1

/product="P27.9"

/protein\_id="AAC61738.1"

/db\_xref="PID:g3661517"

/db\_xref="GI:3661517"

/translation="MTYPRRYRRRRHRRSHLQILRRPWLPHRRHYRWRKNGI  
FNTLSRTFGYTVKRTVTTPSWAVDMRKNIDFVPPGGTGNKISIPFYYIRKVK  
VEFWPCSPITQDGRGVGSTAVILDNFTKATALTYPVYNSRHTIPOPFSYHSRY  
FTPKPVLDSIDYFQPNKRNQLWRLQTSNDVHVLGTAFENSKYDQDYNIRVTMY  
VQREFNLKDPPLP"

BASE COUNT 463 a 357 c 482 g 466 t

ORIGIN

alignment\_scores:

Quality: 572.00 Length: 104

Ratio: 5.500 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-209-961-9 x AF086834/rev ..

Align seg 1/1 to reverse of: AF086834 from: 1 to: 1768

1 MetValThrIleProLeuValPheArgTTPPheProValCysGlyPh 17

|||||

671 ATGGTAACCATCCACCACCTTGTGTTTAGTGGTTTCCAGTATGTTT 622

|||||

17 eArgValCysLysIleSerSerProPheAlaPheThrProArgTTP 34

|||||

621 CCGGGTTTGCAAAATAGCAGCCATTTGCTTTTACCACACCCAGGTGC 572

|||||

34 roHisAsnGluValTyrIleGlyPheProIleThrLeuLeuHisPhePro 50

|||||

571 CCACAATGAAGTGTACATTTGCTTTTCCCAATCAGCTTTTGCATTTTCCC 522

|||||

51 AlaHisPheGlnLysPheSerGlnProAlaGluIlePheAspLysArgTy 67

|||||

521 GCTCACTTTCAAAAGTTCAGCAGCCCGCGGAAATTTTGACAAAGTTA 472

|||||

67 rArgValLeuLeuCysAsnGlyHisGlnAsnProAlaLeuGlnGlyT 84

|||||

471 CAGGGTGTGCTTTTGACAGGTGCTGCTGCTTTCAGATCCCGCTCTCCAACAGTTA 422

```

84 hrHisSerSerArgGlnValThrProLeuSerLeuArgSerArgSer 100
|||||
421 CTCACAGTAGACAGTCACTCCGTTGCTTGGAGATCGAGGAGCTCC 372
|||||
101 ThrPheAsnLys 104
|||||
371 ACATTCAATAAG 360
|||||
seq_name: gb_vi:AF086835

seq_documentation_block:
LOCUS AF086835 1768 bp DNA circular VRL 29-SEP-1998
DEFINITION Porcine circovirus strain 9741, complete genome.
ACCESSION AF086835
NID 93661518
VERSION AF086835.1 GI:3661518
KEYWORDS
SOURCE
ORGANISM
porcine circovirus.
VIRUSES; ssDNA viruses; Circoviridae; Circovirus.
REFERENCE
1 (bases 1 to 1768)
AUTHORS Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.
TITLE Emergence of a new porcine circovirus
JOURNAL Unpublished
AUTHORS Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.
TITLE Direct Submission
JOURNAL Submitted (26-AUG-1998) VIDO, University of Saskatchewan, 120
Veterinary Road, Saskatoon, Saskatchewan S7N 5E3, Canada
FEATURES
source
1. .1768
/organism="porcine circovirus"
/strain="9741"
/db_xref="taxon:46221"
51. .995
/codon_start=1
/product="p35.8"
/protein_id="AAC61739.1"
/db_xref="PIR:G3661519"
/translation="MPSKNGRSGQPQHRKRWFTLNNPSENKRKKIRELPISLFDYFI
VGEENEGRTPHLOGFANFVKQTNKVKVLGARCHIEKAKGTDOQNEKVCSEGN
LLIEGAPRSQGSQSDLSAVSTLLESGLITVAKQHPVTFVKNFGLAELLKVSQKM
QRDRKTNVHFIVGPPGCKSKWANFANPETTYWKPKNKWDGKHGKVVVIDFY
GWLPMDDLRLCDRPLTVTKRGTVPFILARSILITSNQTPLEWYSSTAVPAVEALYR
RITSLVFNKNATQSTEEGGQFVTLSPCCPEFFPEINY"
complement(1034..1735)
/codon_start=1
/product="p27.9"
/protein_id="AAC61740.1"
/db_xref="PIR:G3661520"
/db_xref="GI:3661520"
/translation="MTYPRRYRRRRHRRPSHLGQILRRRPLVHPRHRYRWRKNGI
FNTRLRFTGYTVKRTVTTPSWAVDMREKIDDFVPPGSGTNKISIPFEYVIRKVK
VEFWPCSPITQDGRGVGSFVILDDNFYKATALYDPYNTSSRHTIPQPSYHSRY
FYPKPVLDSTIDYFQPNRRNQLRLQTSNGNVHVLGTFAPENSKIDQDINIRVTMI
VOFREFNLDKPLPK"
CDS
BASE COUNT 464 a 360 c 480 g 464 t
ORIGIN

alignment_scores:
Quality: 572.00 Length: 104
Ratio: 5.500 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-209-961-9 x AF086835/rev ..
Align seg 1/1 to reverse of: AF086835 from: 1 to: 1768
1 MetValThrIleProProLeuValPheArgTrpPheProValCysGlyPh 17

```

```

|||||
671 ATGGTAACCATCCACCACTTCTTTTAGTGGTTTCCAGTATGTGGTTT 622
|||||
17 eargValCysLysIleSerSerProphealapherThrProArgTrpP 34
|||||
621 CCGGGTTTGCAAAATAGCAGCCCACTTCTTTTACCACCCAGGTGGC 572
|||||
34 rOHISAsnGluValTyrIleGlyPheProIleThrLeuLeuHisPhePro 50
|||||
571 CCCACAATGAAGTGTACATTTGTTTCCATCAGCTTTTGCAATTTCCC 522
|||||
51 AlaHisPheGlnLysPheSerGlnProAlaGluIlePheAspLysArgTy 67
|||||
521 GTCACACTTCAAAAGTTCAGCCAGCCCGGAAATTTTGCACAAACGTTA 472
|||||
67 rArgValLeuLeuCysAsnGlyHisGlnAsnProAlaLeuGlnGlyT 84
|||||
471 CAGGGTGTCTTTGCAACGGTCCACAGAAATCCCGCTCTCCAAACAGTA 422
|||||
84 hrHisSerSerArgGlnValThrProLeuSerLeuArgSerArgSer 100
|||||
421 CTCACAGTAGACAGTCACTCCGTTGCTTGGAGATCGAGGAGCTCC 372
|||||
101 ThrPheAsnLys 104
|||||
371 ACATTCAATAAG 360
|||||
seq_name: gb_vi:AF086836

seq_documentation_block:
LOCUS AF086836 1768 bp DNA circular VRL 29-SEP-1998
DEFINITION Porcine circovirus strain M226, complete genome.
ACCESSION AF086836
NID 93661521
VERSION AF086836.1 GI:3661521
KEYWORDS
SOURCE
ORGANISM
porcine circovirus.
VIRUSES; ssDNA viruses; Circoviridae; Circovirus.
REFERENCE
1 (bases 1 to 1768)
AUTHORS Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.
TITLE Emergence of a new porcine circovirus
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 1768)
AUTHORS Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.
TITLE Direct Submission
JOURNAL Submitted (26-AUG-1998) VIDO, University of Saskatchewan, 120
Veterinary Road, Saskatoon, Saskatchewan S7N 5E3, Canada
FEATURES
source
1. .1768
/organism="porcine circovirus"
/strain="M226"
/db_xref="taxon:46221"
51. .995
/codon_start=1
/product="p35.8"
/protein_id="AAC61741.1"
/db_xref="PIR:G3661522"
/db_xref="GI:3661522"
/translation="MPSKNGRSGQPQHRKRWFTLNNPSEDERKKIRELPISLFDYFI
VGEENEGRTPHLOGFANFVKQTNKVKVLGARCHIEKAKGTDOQNEKVCSEGN
LLIEGAPRSQGSQSDLSAVSTLLESGLITVAKQHPVTFVKNFGLAELLKVSQKM
QRDRKTNVHFIVGPPGCKSKWANFANPETTYWKPKNKWDGKHGKVVVIDFY
GWLPMDDLRLCDRPLTVTKRGTVPFILARSILITSNQTPLEWYSSTAVPAVEALYR
RITSLVFNKNATQSTEEGGQFVTLSPCCPEFFPEINY"
complement(1034..1735)
/codon_start=1
/product="p27.9"
/protein_id="AAC61742.1"
/db_xref="PIR:G3661523"
/db_xref="GI:3661523"
/translation="MTYPRRYRRRRHRRPSHLGQILRRRPLVHPRHRYRWRKNGI
FNTRLRFTGYTVKRTVTTPSWAVDMREKIDDFVPPGSGTNKISIPFEYVIRKVK

```

VEFWPCSPITQGRGVSTAVILDNEVTKATALYDYPVYNYSSRHTIPQPFYSHSRY  
FTKPLVLDSTIDYFQPNKNQWLRLQTSNVDHVLGTLAFENSKYDQDYNIRVTWY  
VQFRENLKDPLEP"

BASE COUNT 463 a 363 c 480 g 462 t  
ORIGIN

alignment\_scores:  
Quality: 572.00 Length: 104  
Ratio: 5.500 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-209-961-9 x AF086836/rev ..

Align seg 1/1 to reverse of: AF086836 from: 1 to: 1768

1 MetValThrIleProProLeuValPheArgTrpPheProValCysGlyPh 17  
|||||  
671 ATGGTAACCATCCACCACCTGTTTGTAGTGGTTTCCAGTATGTGTTT 622

17 eArgValCysLysIleSerSerProPheAlaPheThrProArgTrpP 34  
|||||  
621 CGGGGTTTGC AAAATAGCAGCCCATTTGCTTTTACACACCCAGGTGC 572

34 toHisAsnGluValTyrlleGlyPheProIleThrLeuLeuHisPhePro 50  
|||||  
571 CCACAAAGAAAGTACATGTGTTTCCAAATCAGCGTTTGTGATTTTCCC 522

51 AlaHisPheGlnLysPheSerGlnProAlaGluIlePheAspLysArgTy 67  
|||||  
521 GCTCACTTTCAAAAGTTACAGCCAGCCGCGGAAATTTTGACAAACGTTA 472

67 rArgValLeuLeuCysAsnGlyHisGlnAsnProAlaLeuGlnGlyT 84  
|||||  
471 CAGGTTGTCTTTGCACAGGTACACAGAAATCCCGCTCTCCAACAGGTA 422

84 hrHisSerSerArgGlnValThrProLeuSerLeuArgSerArgSer 100  
|||||  
421 CTCACAGCAGTAGACAGGTCACTCCGTTGCTTGTAGATCGAGGAGCTCC 372

101 ThrPheAsnLys 104

371 ACATTCAATAAG 360

seq\_name: gb\_vi:AF118095

seq\_documentation\_block:

LOCUS AF118095 597 bp DNA VRL 27-JUL-1999  
DEFINITION Porcine circovirus type 2 isolate IAF-614 unknown gene.

ACCESSION AF118095

NID 95596425

VERSION AF118095.1 GI:5596425

KEYWORDS

SOURCE

porcine circovirus type 2.

porcine circovirus type 2

Viruses; ssDNA viruses; Circoviridae; Circovirus.

1 (bases 1 to 597)

Quardani,M., Wilson,L., Jette,R., Montpetit,C. and Dea,S.

Multiplex PCR for detection and typing of porcine circoviruses

J. Clin. Microbiol. (1999) In press

2 (bases 1 to 597)

Quardani,M., Wilson,L., Jette,R., Montpetit,C. and Dea,S.

Direct Submission

Submitted (06-JAN-1999) CRMB, INRS-IAF, 531, Boulevard des

Prairies, Laval, Que H7N 4Z3, Canada

Location/Qualifiers

1. .597

/organism="porcine circovirus type 2"

/isolate="IAF-614"

/db\_xref="taxon:85708"

/note="virus isolated from lung, lymph node, spleen and

tonsil tissue from pigs affected by postweaning

CDS

multisystemic wasting syndrome"  
<1..>597

/note="ORF1"

/codon\_start=1

/product="unknown"

/protein\_id="AAD45580.1"

/db\_xref="PID:g5596426"

/db\_xref="GI:5596426"

/translation="SEDERKKIRLEPISLDYFIVGEENEGRTPHLOGFANFVKQ  
TFNKVYVLGARCHIEKAKGTDHONKEYCSRECNLLIECGAPRSQQRSDLSAVSTL  
LESGLVTVAQHPVTFVNRFGLAELLKVGSKMRKDWKNTNHHVIVGPGCGKSKWA  
ANFADPTTYKQPPKRNKMDGYHGEEVVDDFYGLPW"

BASE COUNT 169 a 112 c 170 g 146 t

ORIGIN

alignment\_scores:

Quality: 541.00 Length: 104

Ratio: 5.304 Gaps: 0

Percent Similarity: 98.077 Percent Identity: 95.192

alignment\_block:

US-09-209-961-9 x AF118095/rev ..

Align seg 1/1 to reverse of: AF118095 from: 1 to: 597

1 MetValThrIleProProLeuValPheArgTrpPheProValCysGlyPh 17  
|||||  
549 ATGGTAACCATCCACCACCTGTTTCTAGTGGTTTCCAGTATGTGTTT 500

17 eArgValCysLysIleSerSerProPheAlaPheThrProArgTrpP 34  
|||||  
499 CCGGGTCTCGAAAATTAGCAGCCCATTTGCTTTTACACACCCAGGTGC 450

34 roHisAsnGluValTyrlleGlyPheProIleThrLeuLeuHisPhePro 50  
|||||  
449 CCACAAATGACGTGTACATGGTCTTCCAAATCAGCTTCTGCATTTTCCC 400

51 AlaHisPheGlnLysPheSerGlnProAlaGluIlePheAspLysArgTy 67  
|||||  
399 GCTCACTTTCAAAAGTTACAGCCAGCCGCGGAAATTTTCTGACAAACGTTA 350

67 rArgValLeuLeuCysAsnGlyHisGlnAsnProAlaLeuGlnGlyT 84  
|||||  
349 CAGGTGCTGCTCTGCAACGTCACACAGATCCCGCTCTCCAACAGGTA 300

84 hrHisSerSerArgGlnValThrProLeuSerLeuArgSerArgSer 100  
|||||  
299 CTCACAGCAGTAGACAGGTCACTCCGTTGCTTGTAGATCGAGGAGCTCC 250

101 ThrPheAsnLys 104

249 ACATTCAATAAG 238

seq\_name: gb\_vi:AF027217

seq\_documentation\_block:

LOCUS AF027217 1768 bp DNA circular VRL 14-MAY-1998  
DEFINITION Porcine circovirus strain pmws PCV, complete genome.

ACCESSION AF027217

NID 92689645

VERSION AF027217.1 GI:2689645

KEYWORDS

porcine circovirus.

porcine circovirus

Viruses; ssDNA viruses; Circoviridae; Circovirus.

1 (bases 1 to 1768)

Hamei,A.L., Lin,L.L. and Nayar,G.P.

Nucleotide sequence of porcine circovirus associated with

postweaning multisystemic wasting syndrome in pigs

J. Virol. 72 (6), 5262-5267 (1998)

JOURNAL 98241772

REFERENCE 2 (bases 1 to 1768)



# AUTHORS

Hamel, A.L., Lin, L.L. and Nayar, G.P.S.

## TITLE

Direct Submission  
Submitted (26-SEP-1997) Virology Laboratory, Veterinary Services  
Branch, Manitoba Agriculture, 545 University Crescent, Winnipeg,  
Manitoba R3T 5S6, Canada

## FEATURES

### source

1. .1768  
/location/Qualifiers  
/organism="porcine circovirus"

/strain="pmws PCV"

/db\_xref="taxon:46221"

/note="both strands of seven overlapping PCR fragments  
were sequenced; virus isolated from lung, lymph node,  
spleen and tonsil tissue from pigs affected by post  
weaning multisystemic wasting syndrome"

51. .995

/note="ORF1: similar to Rep protein encoded by

non-pathogenic PCV, GenBank Accession Number U49186;

predicted 35.8 kDa protein"

/codon\_start=1

/product="putative Rep protein"

/protein\_id="AAC59462.1"

/db\_xref="PID:g2689646"

/db\_xref="GI:2689646"

/translation="MPSKNGRSGPQPHKRWFTLNPSDEKKIRELPISLDFYFI  
VGEENEGRTPHLOGFANFVKQTKVYKILGARYIEKAGTDQONKEYCKSGKN  
LITECAPRSQGSQSLSTAVSTLLESGSLTVAEQHPVTFVRFNGLAEVLKVSQGM  
OKRDWNTVHVIVGPPCGKSKWAANFADPTTYWPKPRNKWDGYPGEVVDIFY  
GWLPPDILLRLCDRIPLTVETKGTVPFLARSLITSNQTPLEWYSSTAVPAVEALYR  
RTISLVFNKATEQSTEEGGQFVTLSPCPPEFFYEINY"

117. .125

/note="glycosylation site"

### misc\_feature

327. .332

/note="ORF3: predicted 11.9 kDa protein"

/codon\_start=1

/protein\_id="AAC59464.1"

/db\_xref="PID:g2689648"

/db\_xref="GI:2689648"

/translation="MTIPLVSRWFPVCGFRVCKYISSPFAFTTPRPNDVYIGLPI  
TLHFFPAHFQKFSQPAEISDKYRVLICNGHOTPALQQGTHSRQVTPLSRRSSTF  
NK"

complement(386. .565)

/note="ORF4: predicted 6.5 kDa protein"

/codon\_start=1

/protein\_id="AAC59465.1"

/db\_xref="PID:g2689649"

/db\_xref="GI:2689649"

/translation="MTCTLVFQSRFCIFPLTFKSSASPRKFLTNVTGCCSATVTRLPL  
SNKVLTVDRSLRCP"

complement(470. .478)

/note="glycosylation site"

complement(688. .753)

/note="ORF8: predicted 2.3 kDa protein"

/codon\_start=1

/protein\_id="AAC59469.1"

/db\_xref="PID:g2689653"

/db\_xref="GI:2689653"

/translation="MDIDHTVSVDHPTAAASHKSHQ"

816. .824

/note="glycosylation site"

906. .914

/note="glycosylation site"

983. .988

/note="ORF11: predicted 1.8 kDa protein"

/codon\_start=1

/protein\_id="AAC59472.1"

/db\_xref="PID:g2689656"

/db\_xref="GI:2689656"

/translation="MKKNHYEIKKIQ"

1016. .1177

/note="ORF5: predicted 6.2 kDa protein"

/codon\_start=1

/protein\_id="AAC59466.1"  
/db\_xref="PID:g2689650"  
/db\_xref="GI:2689650"  
/translation="MVFIFHLGFKWGVKIKFSELYIHGYTDIVLVVVTYFERSAEA  
YVHISRGL"

### polyA\_site

#### CDS

complement(1022. .1027)

complement(1034. .1735)

/note="ORF2: predicted 27.8 kDa protein"

/codon\_start=1

/protein\_id="AAC59463.1"

/db\_xref="PID:g2689647"

/db\_xref="GI:2689647"

/translation="MTYPRRYRRRRHRRSHLQILRRPWLHPRHRYWRKNGI  
FNRLSRTFGYTKATVTPSWAVDMRFRNIDDFVPPGGTGNKISIPFYYRIKVK  
VEFWPCSPITQDGRGVGSTAVILDNFTVKATALTYDPVNVSSRHTIPOPEYSYRY  
ETKPVLDSTIDYFQPNKRTQLWLRLQTSRNVHVLGLTAFENSIIQODNIRVTMY  
VQRFNKLQPPPLRP"

complement(1301. .1309)

/note="glycosylation site"

complement(1522. .1611)

/note="ORF6: predicted 3.1 kDa protein"

/codon\_start=1

/protein\_id="AAC59467.1"

/db\_xref="PID:g2689651"

/db\_xref="GI:2689651"

/translation="MASSTPASPAPSDILSRLOSERPPGRWT"

1524. .1631

/note="ORF10: predicted 4.1 kDa protein"

/codon\_start=1

/protein\_id="AAC59471.1"

/db\_xref="PID:g2689655"

/db\_xref="GI:2689655"

/translation="MSTAQEGVLTVVTPVKVRRVRLKMPFELLOR"

1682. .1741

/note="ORF7: predicted 1.9 kDa protein"

/codon\_start=1

/protein\_id="AAC59468.1"

/db\_xref="PID:g2689652"

/db\_xref="GI:2689652"

/translation="MAAGAVSSSAVTPPWIRHS"

complement(join(1732. .1768, 1. .92))

/note="ORF9: predicted 4.6 kDa protein"

/codon\_start=1

/protein\_id="AAC59470.1"

/db\_xref="PID:g2689654"

/db\_xref="GI:2689654"

/translation="MWLGSSASSILLAGHVAEVLPRCCRSALVILTAHFFRFOL"

join(1746. .1768, 1. .13)

/function="putative replication site"

join(1762. .1768, 1. .2)

/note="AAGTATAC" is similar to the nonanucleotide motif  
in the non-pathogenic PCV, GenBank Accession Number  
U49186"

BASE COUNT 452 a 360 c 495 g 461 t

### ORIGIN

### alignment\_scores:

Quality: 541.00 Length: 104  
Ratio: 5.304 Gaps: 0  
Percent Similarity: 98.077 Percent Identity: 95.192

### alignment\_block:

US-09-209-961-9 x AF027217/rev ..

Align seg 1/1 to reverse of: AF027217 from: 1 to: 1768

1 MetValThrIleProPheValPheArgTrpPheProValCysGlyPh 17

|||||  
671 ATGGTAACCATCCACCACTGTTTCTAGTGTGTTCCAGTATGTGTTT 622

.. 17 eargValCysLysIleSerProphealaPheThrThrProArgTrp 34

|||||

```

621 CCGGTCGCAAAATTAGCAGCCCAATTGCTTTTACCACACCCAGGTGGC 572
34 rohisasnGluValThrIleGlyPheProIleThrLeuHisPhePro 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
571 CCACAAATGACGTGATGCTTCCCAATCAGCGCTCGCATTTTCCC 522
51 AlaHisPheGlnLysPheSerGlnProAlaGluIlePheAspLysArgTy 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
521 GCTCACTTCAAAAGTTCACCGAGCGCGGGAATTTCTGACAAAGCTTA 472
67 rArgValLeuLeuCysAsnGlyHisGlnAsnProAlaLeuGlnGlnGlyT 84
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
471 CAGGTGCTGCTGCAACGGTCACAGACTCCGCTCTCCAAACAGGTA 422
84 hrHisSerSerArgGlnValThrProLeuSerLeuArgSerArgSerSer 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
421 CTCACAGCAGTACAGAGTCACTCCGTTGCTTGAGATCGAGGAGTCC 372
101 ThrPheAsnLys 104
|||||:|||||
371 ACATTCAATAAG 360
seq_name: gb_vi:AF055391
seq_documentation_block:
LOCUS AF055391 1768 bp DNA circular VRL 13-SEP-1998
DEFINITION Porcine circovirus type II from USA, complete genome.
ACCESSION AF055391
NID 93598796
VERSION AF055391.1 GI:3598796
KEYWORDS
SOURCE porcine circovirus.
ORGANISM
Viruses; ssDNA viruses; Circoviridae; Circovirus.
REFERENCE
1 (bases 1 to 1768)
Meehan,B.M., McNelly,F., Todd,D., Kennedy,S., Jewhurst,V.A.,
Ellis,J.A., Hassard,L.E., Clark,E.G., Haines,D.M. and Allan,G.M.
Characterization of novel circovirus DNAs associated with wasting
syndromes in pigs
J. Gen. Virol. 79 (Pt 9), 2171-2179 (1998)
JOURNAL
MEDLINE 98418498
REFERENCE 2 (bases 1 to 1768)
Meehan,B.M.
AUTHORS
Direct Submission
TITLE
Submitted (26-MAR-1998) Brian M Meehan, Veterinary Science, The
Queen's University of, Belfast, Stoney Road, Belfast, Northern
Ireland, BT4 3SD, U.K.
FEATURES
Location/Qualifiers
1..1768
/organism="porcine circovirus"
/specific_host="Sus scrofa"
/strain="Porcine circovirus Type II"
/db_xref="taxon:46221"
/clone="Imp.999 Ecoli No.8"
/country="USA"
complement(101..190)
CDS
/codon_start=1
/product="ORF6"
/protein_id="AAC35308.1"
/db_xref="PID:g3598807"
/db_xref="GI:3598807"
/translation="MASSTPASPAPSDILRLPQSERPPGRWT"
103..210
CDS
/codon_start=1
/product="ORF10"
/protein_id="AAC35306.1"
/db_xref="PID:g3598805"
/db_xref="GI:3598805"
/translation="MSTAQEGVLTVVALTVPVKVRRRLKMPFFLLQR"
261..320
CDS
/codon_start=1
/product="ORF7"
/protein_id="AAC35307.1"

```

```

/db_xref="GI:3598806"
/db_xref="GI:3598806"
/translation="MAAGAVSSSAVTPPWIRHS"
complement(311..439)
/notes="predicted 4.6 kDa protein"
/codon_start=1
/product="ORF9"
/protein_id="AAC35304.1"
/db_xref="PID:g3598803"
/db_xref="GI:3598803"
/translation="MWLGSSAILLAGHVAEEVLPCCRCRSALVILTAHFFRQL"
398..1342
CDS
/codon_start=1
/product="putative Rep protein (ORF1)"
/protein_id="AAC35298.1"
/db_xref="PID:g3598797"
/db_xref="GI:3598797"
/translation="MPSKNGRSGQPQHKRWVFTLNNPSEDERKKIRELPISLDFYFI
VGEENEGRTPLHGFANFVKQTFNKVYLGAICHIEKAGTQONKEYCSKEGN
LLIEGAPRSQGSRLSTAVSTLLESGLVTVAEOHPVTFVNFRLGLAELLKVSQKM
QKRDNKTNVHVIIVGPPCGCKSKWAANFADPTIYWKPPRNKWDGHHGEVVIDDFY
GWLPDLDRLCDRIPLIVETIKGIVPFLARSLIITSNQTPLEWISSIAVFAVALTR
RITSLVFWKNATEQSTEEGGQFVTLSPPCPEPEYIN"
574..679
polyA_signal
CDS
complement(704..1018)
/notes="predicted 11.9 kDa protein"
/codon_start=1
/product="ORF3"
/protein_id="AAC35300.1"
/db_xref="PID:g3598799"
/db_xref="GI:3598799"
/translation="MVTIPLVSRWFPVCGFRVCKISSPFAFTTPRPHNDVYIGLPI
TILHFPAPHQKFSQPAEISDKRYRVLCLNGHQTALQOQTHSSRQVTPLSRSTTF
NK"
complement(733..912)
CDS
/notes="predicted 6.5 kDa protein"
/codon_start=1
/product="ORF4"
/protein_id="AAC35301.1"
/db_xref="PID:g3598800"
/db_xref="GI:3598800"
/translation="MCTCLVFSQRCIFPLTFKSSAPRRKFLTNVTGCCSATVTRLPL
SNKVLTAVDRLRCP"
complement(1035..1100)
CDS
/notes="predicted 2.3 kDa protein"
/codon_start=1
/product="ORF8"
/protein_id="AAC35303.1"
/db_xref="PID:g3598802"
/db_xref="GI:3598802"
/translation="MDIDHTVSDHPTAASHKSHQ"
1330..1335
polyA_signal
CDS
complement(1336..1380)
/notes="predicted 1.8 kDa protein"
/codon_start=1
/product="ORF11"
/protein_id="AAC35305.1"
/db_xref="PID:g3598804"
/db_xref="GI:3598804"
/translation="MNNKNHYEVKKTQ"
1363..1524
CDS
/notes="predicted 6.2 kDa protein"
/codon_start=1
/product="ORF5"
/protein_id="AAC35302.1"
/db_xref="PID:g3598801"
/db_xref="GI:3598801"
/translation="MVFIIHLGFKWGVKIKFSELYIHGTYDVLVYVYVFSASA
YVYHISRGL"
complement(1369..1374)
polyA_site
complement(join(1381..1768,1..314))
CDS
/notes="predicted 27.8 kDa protein"
/codon_start=1

```



```
/note="predicted 2.3 kDa protein"
/codon_start=1
/product="ORF8"
/protein_id="AAC35314.1"
/db_xref="PID:g3598814"
/db_xref="GI:3598814"
/translation="MDIDHTVSDVHDPTAAASHKSHQ"
1330..1335
complement(1336..1380)
/note="predicted 1.8 kDa protein"
/codon_start=1
/product="ORF11"
/protein_id="AAC35316.1"
/db_xref="PID:g3598816"
/db_xref="GI:3598816"
/translation="MNNKNHYEVKKQ"
1363..1524
/note="predicted 6.2 kDa protein"
/codon_start=1
/product="ORF5"
/protein_id="AAC35313.1"
/db_xref="PID:g3598813"
/db_xref="GI:3598813"
/translation="MVFIHLGFKGVFKIKFSELYINGYTDIVLVVVTTFERSAEA
VVHISRGL"
complement(1369..1374)
/note="predicted 27.8 kDa protein"
/codon_start=1
/product="ORF2"
/protein_id="AAC35310.1"
/db_xref="PID:g3598810"
/db_xref="GI:3598810"
/translation="MTTPRRYRRRRHPRSHLQGLRRRLVHPHRYRWRKNGI
FNTLSRTFTGTVARTVTTPSWADMMRFKIDDFVPGGNTKISIPFEYIRKVK
VEFWPCPTIGDGRGVSTAVLDDNFVTKATLTDYDYNYSRHTIPOPFYSHSRY
FTPKPVLDSIDYDQPNKRNQLRLQTSNGVDHVLGAAFENSKYDQDYNIRVTMY
VQPFENLKDPLKP"
BASE COUNT 451 a 361 c 495 g 461 t
ORIGIN

alignment_scores:
  Quality: 541.00 Length: 104
  Ratio: 5.304 Gaps: 0
Percent Similarity: 98.077 Percent Identity: 95.192

alignment_block:
US-09-209-961-9 x AF055392/rev ...
Align seg 1/1 to reverse of: AF055392 from: 1 to: 1768

1 MetValThrLeuProLeuValPheArgTrpPheProValCysGlyP 17
|||||
1018 ATGTAACCATCCACCACCTGTTCTAGTGGTGTTCACAGTGTGTT 969
|||||

17 eargValCysLysIleSerSerProPheAlaPheThrThrProArgTrp 34
|||||
968 CCGGGTCTGCAAAATAGCAGCCATTTGCTTTTACACACCCAGGTGC 919
|||||

34 roHisAsnGluValTyrIleGlyPheProIleThrLeuLeuHisPhePro 50
|||||
918 CCACAATGACGGTGCATATGGTCTTCCAAATCAGCTTCGCAATTTC 869
|||||

51 AlaHisPheGlnLysPheSerGlnProAlaGluIlePheAspLysArgTy 67
|||||
868 GCTCACTTTCAAAAGTTACAGCAGCCCGGGAATTTCTGACAAACGTTA 819
|||||

67 rArgValLeuLeuCysAsnGlyHisGlnAsnProAlaLeuGlnGlnGly 84
|||||
818 CAGGTGCTGCTGCAACGGTCACACAGACTCCCGCTCTCCAAACAGGTA 769
|||||

84 hrHisSerArgGlnValThrProLeuSerLeuArgSerArgSer 100
|||||

|||||
768 CTCACAGCAGTAGACAGGTCACTCCGTTGTCTTGAGATCGAGAGCTCC 719
|||||
101 ThrPheAsnLys 104
|||||
718 ACATTCAATAAG 707

seq_name: gb_v1:AF109397
seq_documentation_block:
LOCUS AF109397 1768 bp DNA circular VRL 02-FEB-1999
DEFINITION Bovine circovirus, complete genome.
ACCESSION AF109397
NID 94210313
VERSION AF109397.1 GI:4210313
KEYWORDS
SOURCE
ORGANISM
bovine circovirus.
Viruses; ssDNA viruses; Circoviridae; Circovirus.
REFERENCE
1 (bases 1 to 1768)
AUTHORS Hamel,A.L. and Nayar,G.P.S.
TITLE Nucleotide sequence of a circovirus detected in cattle with various
clinical syndromes
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 1768)
AUTHORS Hamel,A.L. and Nayar,G.P.S.
TITLE Direct Submission
JOURNAL Submitted (27-NOV-1998) Virology Laboratory, Veterinary Diagnostic
Laboratory, Manitoba Agriculture, 545 University Crescent,
Winnipeg, Manitoba R3T 5S6, Canada
REFERENCE
3 (bases 1 to 1768)
AUTHORS Hamel,A.L. and Nayar,G.P.S.
TITLE Direct Submission
JOURNAL Submitted (02-FEB-1999) Virology Laboratory, Veterinary Diagnostic
Laboratory, Manitoba Agriculture, 545 University Crescent,
Winnipeg, Manitoba R3T 5S6, Canada
REMARK Sequence update by submitter
COMMENT On Feb 2, 1999 this sequence version replaced gi:4028609.
FEATURES
location/Qualifiers
1..1768
/organism="bovine circovirus"
/virus
/db_xref="taxon:85542"
/note="BCV"
sequence was obtained from several overlapping PCRs using
DNA extracted from various tissues from cattle; similar to
Porcine circovirus sequence presented in GenBank Accession
Number AF027217"
13..35
/repeat_region
13..35
/rpt_type=tandem
/rpt_unit=13..18
51..995
CDS
/note="ORF-1; similar to Porcine circovirus ORF1 encoded
by the sequence presented in GenBank Accession Number
AF027217."
/codon_start=1
/product="putative Rep and coat protein"
/protein_id="AAD11928.1"
/db_xref="PID:g4210314"
/db_xref="GI:4210314"
/translation="MPSKNGRSGPQPKRWVFTLNNPSEDERKKIRLPISLFDYFI
VGEENEGRPHLQGFANFYKQTFNKVWLGARCHEKAKGTDOQNKYCSKRGK
LLIEGAPRSQGSRLSTAVSLTLESGLVFAEOHPVTFVFNFGLAELKLVSGKM
QKRWKTNHVIVGPPGCGSKAAAFADPTTYKPKRNKWDGSHGEVVYIDDFY
GWLPMDDLRLCDRPLVETGKGVFPFLARSILITSNQPLEWYSSTAVPAVEALYR
RITSLVFWKNATEQSTEEGGQFVTLSPPCPEPEYIN"
327..332
complement(357..671)
/note="similar to Porcine circovirus ORF-3 encoded by the
sequence presented in GenBank Accession Number AF027217."
/codon_start=1
/product="ORF-3"
/protein_id="AAD11930.1"
polya_signal
CDS
```

```

/db_xref="PID:g4210315"
/db_xref="GI:4210315"
/translation="MTIPPLVSRWPPVCGFRVCKISSPPAFTPRPWPHNDVYIGLPI
TLHFPAPHQKFSQPAEISDKRYVLLNGHQTPALQQGTHSSRQVTPLSLRSSSTF
NK"
CDS
complement(386..565)
/notes="similar to Porcine circovirus ORF-4 encoded by the
sequence presented in GenBank Accession Number AF027217."
/codon_start=1
/product="ORF-4"
/protein_id="AADI1931.1"
/db_xref="PID:g4210316"
/db_xref="GI:4210316"
/translation="MTCTLVFQSRFCIFPLTEKSSAPRKFLNVTGCCSATVTRLPL
SNKLVAVDRSLRCP"
CDS
553..732
/codon_start=1
/product="ORF-12"
/protein_id="AADI1927.1"
/db_xref="PID:g4028610"
/db_xref="GI:4028610"
/translation="MTSLWGLGVVYKANGLLILQTRKPKHTGNHLETSGGMVTVKWK
LLMTFACRGMY"
CDS
complement(688..753)
/codon_start=1
/product="ORF-8"
/protein_id="AADI1935.1"
/db_xref="PID:g4028618"
/db_xref="GI:4028618"
/translation="MDIDHTSVSDHPTAASHKSHQ"
983..988
complement(989..1033)
/codon_start=1
/product="ORF-11"
/protein_id="AADI1937.1"
/db_xref="PID:g4028620"
/db_xref="GI:4028620"
/translation="MNKNHYEVIKKTQ"
1016..1177
/codon_start=1
/product="ORF-5"
/protein_id="AADI1932.1"
/db_xref="PID:g4028615"
/db_xref="GI:4028615"
/translation="WVFIHLGKWCVKFKFSELYIHGYTDIVLVVTFVRSAAE
YVWHISRGL"
polyA_signal
complement(1022..1027)
/notes="similar to Porcine circovirus ORF-2 encoded by the
sequence presented in GenBank Accession Number AF027217."
/codon_start=1
/product="ORF-2"
/protein_id="AADI1929.1"
/db_xref="PID:g4028612"
/db_xref="GI:4028612"
/translation="MTYPRRRYRRTRPRSHLQGLRRPVLVHPRHRYWRKNGI
FNRUSRTGYTKATVTPSWAVDMRFNIDDFVPPGGTNNKISPEYYIRKVK
VFVPCSPITQDGRGVGTAVILDDNFVKATALTDPYVYVSSRHTIPQPFYSYHR"
complement(1522..1611)
/codon_start=1
/product="ORF-6"
/protein_id="AADI1933.1"
/db_xref="PID:g4028616"
/db_xref="GI:4028616"
/translation="MASSTPASPAPSDILSRLPQSRPPGRWT"
1524..1631
/codon_start=1
/product="ORF-10"
/protein_id="AADI1936.1"
/db_xref="PID:g4028619"
/db_xref="GI:4028619"
/translation="NSTAQEGVLTVVATVYKVRERRVLKMPFFLLQR"
1682..1741
CDS

```

```

/codon_start=1
/product="ORF-7"
/protein_id="AADI1934.1"
/db_xref="PID:g4028617"
/db_xref="GI:4028617"
/translation="MAAGAGPSSAVTPPWIRHS"
525n(1750..1768,1..13)
join(1762..1768,1..2)
/notes="putative; similar to the nonnucleotide motif of
porcine circovirus presented in GenBank Accession Numbers
AF027217 and U49186"
BASE COUNT 452 a 362 c 493 g 461 t
ORIGIN

alignment_scores:
  Quality: 541.00      Length: 104
  Ratio: 5.304        Gaps: 0
  Percent Similarity: 98.077 Percent Identity: 95.192

alignment_block:
US-09-209-961-9 x AF109397/rev ..

Align seg 1/1 to reverse of: AF109397 from: 1 to: 1768

1 MetValThrIleProLeuValPheArgTTPheProValCysGlyPh 17
|||||
671 ATGGTAACCATCCACCACCTGTTCTAGGTGGTTTCCAGATGTGGTTT 622

17 eArgValCysLysIleSerSerProPheAlaPheThrThrProArgTTP 34
|||||
621 CCGGGCTCGAATAATAGCAGCCCATTTGCTTTACACACCCAGGTGC 572

34 rHISAsnGluValTyrIleGlyPheProIleThrLeuLeuHisPhePro 50
|||||
571 CCCAATGACGCTGATACATTGGTCTTCCAAATCAGCTTCTGCATTTTCC 522

51 AlaHisPheGlnLysPheSerGlnProAlaGluIlePheAspLysArgTy 67
|||||
521 GTCACCTTCAAAAGTTTACGCCAGCCCGGAAATTTCTGCACAAACGTTA 472

67 rArgValLeuLeuCysAsnGlyHisGlnAsnProAlaLeuGlnGlnGlyT 84
|||||
471 CAGGTGCTGCTCTGCAAGGTCACAGACTCCCGCTCTCCACACAGGTA 422

84 hrHisSerSerArgGlnValThrProLeuSerLeuArgSerArgSerSer 100
|||||
421 CTCACAGCAGTAGACAGGTCACCTCGGTTGCTTGTGAGATCGAGGAGCTCC 372

101 ThrPheAsnLys 104
|||||
371 ACATTCAATAAG 360

seq_name: gb_vi:AF147751

seq_documentation_block:
LOCUS AF147751 1768 bp DNA 08-JUN-1999
DEFINITION Porcine ci/c. virus type 2 unknown genes.
ACCESSION AF147751
NID 95007010
VERSION AF147751.1 GI:5007010
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 1768)
AUTHORS Pogranichnyy,R., Yoon,K.-J., Harms,P., Swenson,S., Zimmerman,J. and
Sorden,S.
TITLE Characterization of clinical and immune responses in young swine to
experimental porcine circovirus type II infection
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1768)
AUTHORS Yoon,K.-J. and Pogranichnyy,R.M.

```

TITLE Direct Submission  
JOURNAL Submitted (30-APR-1999) Veterinary Diagnostic Laboratory, Iowa  
State University, College of Veterinary Medicine, Ames, Iowa 50011,  
USA

## FEATURES

## source

1. .1768 Location/Qualifiers  
/organism="porcine circovirus type 2"  
/isolate="ISUVDL 98-15237"  
/specific\_host="Sus scrofa"  
/db\_xref="taxon:85708"  
/note="isolated from pigs with postweaning multisystemic  
wasting syndrome"  
51. .995  
/note="ORF1; 35-36kD"  
/codon\_start=1  
/product="unknown"  
/protein\_id="AAD37776.1"  
/db\_xref="PID:95007011"  
/db\_xref="GI:5007011"  
/translation="MPSKNGRSGPQPHKRWFTLNNPSEDERKKIRELPISLFDYFI  
VGEENEGRTPLHGFANFKVKTGLGAKHIEKAKGTDOONKEVCSEKGN  
LLIECGAPRSQGRSDLSSTAVSLLESGLTVAEQHPVTFVRNFRGLAELLKVSCKM  
QKRWKTNVHVIVGPPGCKSKWAANFADPTTYWKPKNWMDGIRGEVVVIDDFY  
GWLPMDDLRLCDRYPLTVETGKGVPELARSILITSNOTPLEWYSTAVPAVEALYR  
RITSLVFNKATEQSTEEGGQFVTLSPPCPEFFPEYIN"  
complement(1034. 1735)  
/note="ORF2; 27-28kD"  
/codon\_start=1  
/product="unknown"  
/protein\_id="AAD37777.1"  
/db\_xref="PID:95007012"  
/db\_xref="GI:5007012"  
/translation="MTYPRRYRRRRHRRSHLGOILRRPWLHPHRYWRKNGI  
ENTLSRFGTVATVTFPSWADMMRFNIDFVPPGGTGNKISIFEFYIRKVK  
VEFPSCSITGDRGVGTAVLNDNFVKATALIDYVNFSSRHTIPOPFSCHRY  
FTPKPLDSTIDYFQPNKRNQLRLQTSRNVHDVGLGTAFENSIIYQDYNIRVTY  
VQFREFNKKDPLKP"

## CDS

BASE COUNT 450 a 362 c 494 g 462 t  
ORIGIN  
alignment\_scores:  
Quality: 541.00 Length: 104  
Ratio: 5.304 Gaps: 0  
Percent Similarity: 98.077 Percent Identity: 95.192

## CDS

alignment\_block:  
US-09-209-961-9 x AF147751/rev ..  
Align seg 1/1 to reverse of: AF147751 from: 1 to: 1768

1 MetValThrIleProLeuValPheArgTrpPheProValCysGlyPh 17  
|||||  
671 ATGGTAACCATCCACACACATGTTCTTAGTGTTTCCAGTATGTGTTT 622  
17 eArgValCysLysIleSerProPheAlaPheThrThrProArgTrpP 34  
|||||  
621 CCGGGCTGCAAAATAGACGCCATTTGCTTTTACCACACCCAGGTGGC 572  
34 roHisAsnGluValTyrIleGlyPheProIleThrLeuLeuHisPhePro 50  
|||||  
571 CCCACAATGACGTGATACATTGGCTTCCATCATCAGCTTCGTGCAATTTCC 522  
51 AlaHisPheGlnIysPheSerGlnProAlaGluIlePheAspIysArgTy 67  
|||||  
521 GCTCACTTTCAAAAGTTCAGCACGCCGCGGAAATTTCTGACAAACGTTA 472  
67 rArgValLeuLeuCysAsnGlyHisGlnAsnProAlaLeuGlnGlnGlyT 84  
|||||  
471 CAGGGTGCTGCTCTGCAACGGTCACACAGACTCCCGCTCTCCAAACAGSTA 422  
84 hrHisSerSerArgGlnValThrProLeuSerLeuArgSerArgSer 100  
|||||

421 CTCACAGCAGTAGACAGGTCACTCCGTTGTCTTCAGATCGAGGAGCTCC 372

101 ThrPheAsnLys 104

|||||  
371 ACATTCATAAG 360

seq\_name: gb\_v1:AF109399

seq\_documentation\_block:

LOCUS AF109399 1768 bp DNA circular VRL 06-JAN-1999  
DEFINITION Porcine circovirus type 2-E, complete genome.

ACCESSION AF109399

NID 94106896

VERSION AF109399.1 GI:4106896

## KEYWORDS

SOURCE porcine circovirus type 2-E.

ORGANISM porcine circovirus type 2-E

Viruses; ssDNA viruses; Circoviridae; Circovirus; #69#ine

REFERENCE 1 (bases 1 to 1768)

AUTHORS Hamel, A.L. and Nayar, G.P.S.

TITLE Nucleotide sequence of four different isolates of circovirus

JOURNAL detected in pigs with various clinical syndromes

REFERENCE 2 (bases 1 to 1768)

AUTHORS Hamel, A.L. and Nayar, G.P.S.

TITLE Direct Submission

JOURNAL Submitted (27-NOV-1998)

Location/Qualifiers

1. .1768

/organism="porcine circovirus type 2-E"

/viroion

/db\_xref="taxon:85544"

/note="similar to Porcine circovirus sequence presented in  
GenBank Accession Number AF027217; type-E designation is  
based upon restriction endonuclease digestion pattern;  
sequence obtained from several overlapping PCRs using DNA  
extracted from lung, mesenteric lymph node and tonsil of  
pig"

repeat\_region 13. .35

/rpt\_type=tandem

/rpt\_unit=13. .18

51. .995

/note="ORF-1"

/codon\_start=1

/product="putative Rep and coat protein"

/protein\_id="AAD03071.1"

/db\_xref="PID:94106897"

/db\_xref="GI:4106897"

/translation="MPSKNGRSGPQPHKRWFTLNNPSEDERKKIRELPISLFDYFI  
VGEENEGRTPLHGFANFKVKTGLGAKHIEKAKGTDOONKEVCSEKGN  
LLIECGAPRSQGRSDLSSTAVSLLESGLTVAEQHPVTFVRNFRGLAELLKVSCKM  
QKRWKTNVHVIVGPPGCKSKWAANFADPTTYWKPKNWMDGIRGEVVVIDDFY  
GWLPMDDLRLCDRYPLTVETGKGVPELARSILITSNOTPLEWYSTAVPAVEALYR  
RITSLVFNKATEQSTEEGGQFVTLSPPCPEFFPEYIN"

polyA\_signal

complement(357. .671)

/codon\_start=1

/product="ORF-3"

/protein\_id="AAD03073.1"

/db\_xref="PID:94106899"

/db\_xref="GI:4106899"

/translation="WVTIPLVSRWFPVCGFRVCKISSPFAFTTPRPWPHNDVYIGLPI  
TLHFPAHFQKFSQPAEISDKRYRLLCNGHOTPALQOQTHSSRQVTPLSLRSSSTL  
NK"

complement(386. .565)

/codon\_start=1

/product="ORF-4"

/protein\_id="AAD03074.1"

/db\_xref="PID:94106900"

/db\_xref="GI:4106900"

```
/translation="MTCTLVQSRFCIFPLTFKSSASPRKFLNLTGTCSSATVRLPL
SNKLVAVDRSLRCP"
553..1732
/codon_start=1
/product="ORF-12"
/protein_id="AAD03081.1"
/db_xref="PID:g4106907"
/db_xref="GI:4106907"
/translation="MTSLWHLGVKANGLLILQTRKPTGNHLETSGGMVTWVKW
LLMTFMAGCGMILY"
complement(688..753)
/codon_start=1
/product="ORF-8"
/protein_id="AAD03078.1"
/db_xref="PID:g4106904"
/db_xref="GI:4106904"
/translation="MDIDHVSVDHPTAASHKSHQ"
983..988
complement(989..1033)
/codon_start=1
/product="ORF-11"
/protein_id="AAD03080.1"
/db_xref="PID:g4106906"
/db_xref="GI:4106906"
/translation="MNNKNHYEVIKKTQ"
1016..1363
/notes="similar to Porcine circovirus ORF-5 encoded by the
sequence presented in GenBank Accession Number AF027217."
/codon_start=1
/product="ORF-5"
/protein_id="AAD03075.1"
/db_xref="PID:g4106901"
/db_xref="GI:4106901"
/translation="MFIHLGFKWVKIKFSELYLHGTDIYVLVTFVFSABA
YVHISGLSHLPQIPFVILWLEIVNSGIRKFCPEVPGVGGGLGDCMAGGVYVGV
IGLGCGLYKVVV"
complement(1022..1027)
complement(1034..1735)
/codon_start=1
/product="ORF-2"
/protein_id="AAD03072.1"
/db_xref="PID:g4106898"
/db_xref="GI:4106898"
/translation="MYPRFRERRRHPRLHQLRRPRLVHPHRYRWKKNGL
FNALRSTGYTKATIVSPSWAVDMLRENLDVFPVGGTNKISIPPEYIRKRVK
VEWPCSPITQDGRGVSSAILLDNFVIRKATQYDYPVYNSRHTIPQPFYHSRY
FTPKVLDSTIDYFQPNKRNQLMRLQTSRNDVHGLGTAFENSKYDQDYNIRVTMY
VQFRFNLDKPPPLKP"
1524..1631
/codon_start=1
/product="ORF-10"
/protein_id="AAD03079.1"
/db_xref="PID:g4106905"
/db_xref="GI:4106905"
/translations="MSTAQSGVLTVALTYVYKVRERRALKMPFFLQR"
complement(1528..1611)
/codon_start=1
/product="ORF-6"
/protein_id="AAD03076.1"
/db_xref="PID:g4106902"
/db_xref="GI:4106902"
/translation="MASSTPASPDSILSRPQSRPPGR"
1682..1741
/codon_start=1
/product="ORF-7"
/protein_id="AAD03077.1"
/db_xref="PID:g4106903"
/db_xref="GI:4106903"
/translation="MAAGAVSSSAETPPPIWRHS"
join(1750..1768.1..13)
join(1762..1768.1..2)
/notes="putative; similar to the nonanucleotide motif of
Porcine circovirus presented in GenBank Accession Numbers
```

```
BASE COUNT      448 a      362 c      495 g      463 t
ORIGIN

alignment_scores:
  Quality: 535.00      Length: 104
  Ratio: 5.245      Gaps: 0
  Percent Similarity: 98.077      Percent Identity: 94.231

alignment_block:
US-09-209-961-9 x AF109399/rev ..
Align seg 1/1 to reverse of: AF109399 from: 1 to: 1768

1 MetValThrIleProProLeuValPheArgTrpPheProValCysGlyPh 17
|||||
671 ATGGTAACCATCCACCACTTGTCTTAGTGTTTCCAGTATGTGTTT 622

17 eargValCysLysIleSerSerProPheAlaPheThrThrProArgTrp 34
|||||
621 CCGGGTCTGCAAAATAGCAGCCCATTTGCTTTACACACCCAGGTGGC 572

34 rOHISnGluValTyrlleGlyPheProIleThrLeuLeuHisPhePro 50
|||||
571 CCACAATCAGCTGTACATTTGCTTCCCAATCAGCTTCTGCATTTCCC 522

51 AlaHisPheGlnLysPheSerGlnProAlaGluIlePheAspLysArgTy 67
|||||
521 GCTCACTTTCAAAAGTTCAGCCAGCCGCGGAAATTTCTGACAAAGTTA 472

67 rargValLeuLeuCysAsnGlyHisGlnAsnProAlaLeuGlnGlnGlyT 84
|||||
471 CAGGGTGTGCTCTGCAACGGTCACAGACTCCCGCTCTCCAAACAGGTA 422

84 hrHisSerSerArgGlnValThrProLeuSerLeuArgSerArgSerSer 100
|||||
421 CTCACAGCAGTAGACAGGTCTCCCTTGTCTTGAGTCGAGGAGGTCC 372

101 ThrPheAsnLys 104
|||||
371 ACACCTCAATAAG 360

seq_name: gb_vi:PCAJ3185

seq_documentation_block:
LOCUS      PCAJ3185      1768 bp      DNA      circular      VRL      03-JUL-1998
DEFINITION Porcine circovirus DNA, complete genome, isolate ISU-31.
ACCESSION  AJ223185
NID        93293040
VERSION    AJ223185.1 GI:3293040
KEYWORDS   complete genome.
SOURCE     porcine circovirus.
ORGANISM   porcine circovirus
REFERENCE  1 (bases 1 to 1768)
AUTHORS   Morozov, I.
TITLE     Direct Submission
JOURNAL
REFERENCE  2 (bases 1 to 1768)
AUTHORS   Morozov, I., Sirinarumrit, T., Sorden, S.D., Halbur, P.G., Morgan, M.K.,
Yoon, K.J. and Paul, P.S.
TITLE     Detection of a novel strain of Porcine circovirus in pigs with
Postweaning Multisystemic Wasting Syndrome (PMWS)
JOURNAL   Unpublished
FEATURES   Location/Qualifiers
           1..1768
            /organism="porcine circovirus"
            /virus
            /isolate="ISU-31"
            /specific_host="Sus scrofa"
```

```

CDS
/db_xref="taxon:46221"
51..995
/contig="ORF1"
/codon_start=1
/protein_id="CA11157.1"
/db_xref="PID:ei310034"
/db_xref="PID:g3293041"
/db_xref="GI:3293041"
/translation="MPSKNGSGPQPHKRVFTLNPSDEKRIKELPISLFDYFI
VGEENEGRTPLHGFANFKQFENKVKWLGARCHEKAKGTDOONKEYCSKEGN
LLIEGAPRSQGRSLSSTAVSTLLESGSLVVAHQHPVTFVNRGLAELLKVSQGM
QRDKWNTNHHVIVGPPGCKSKAANFADPTTYWPKRNKWDGHEEVVVDIFY
GWLWDLLRLCDRPLTITGKGVFPFLARSILITSNQTPLEWYSSTAVPAVEALYR
RITSLVFWKNATEQSTEGGQFVTLSPPCPEPEYIN"
BASE COUNT 452 a 361 c 492 g 463 t
ORIGIN

alignment_scores
Quality: 532.00 Length: 104
Ratio: 5.267 Gaps: 0
Percent Similarity: 97.115 Percent Identity: 94.231

alignment_block
US-09-209-961-9 x PCAJ3185/rev ..
Align seg 1/1 to reverse of: PCAJ3185 from: 1 to: 1768
1 MetValThrIleProLeuValPheArgTrpPheProValCysGlyph 17
|||||
671 ATGGTAACCATCCACACACTGTTCTTAGTGTGTTCCAGTATGTTGTT 622

17 eArgValCysIleSerSerProPheAlaPheThrThrProArgTrp 34
|||||
621 CCGGCTGCGAAATTAGACCATTTGCTTTTACCACACCCAGGTGGC 572

34 roHisAsnGluValTyrIleGlyPheProIleThrLeuLeuHisPhePro 50
|||||
571 CCCACATGACGTGTACATTGCTTCCATCAGCTCTGCAATTTCCC 522

51 AlaHisPheGlnIlePheSerGlnProAlaGluIlePheAspIleArgTy 67
|||||
521 GCTCATTTCAAAGTTTCAGTACGCGCGGAAATTTCTGACAAACGTTA 472

67 rArgValLeuLeuCysAsnGlyHisGlnAsnProAlaLeuGlnGlnGlyT 84
|||||
471 CAGGCTGCTCTGCAACGCTCACCAGACTCCGCGCTCTCCAAACAGGTA 422

84 hrHisSerArgGlnValThrProLeuSerLeuArgSerArgSerSer 100
|||||
421 CTCACACAGTAGACAGGTCTACTCCGTGTCCTTGAGATCGAGAGCTCC 372

101 ThrPheAsnIle 104
|||||
371 ACATTCAATAAG 360

seq_name: gb_vi:AF112862
seq_documentation_block:
LOCUS AF112862 1768 bp DNA circular VRL 06-JAN-1999
DEFINITION Porcine circovirus type 2-B, complete genome.
ACCESSION AF112862
NID 94106915
VERSION AF112862.1 GI:4106915
KEYWORDS
SOURCE
ORGANISM porcine circovirus type 2-B.
VIRUSES: ssDNA viruses; Circoviridae; Circovirus;.
REFERENCE 1 (bases 1 to 1768)
AUTHORS Hamel,A.L. and Nayar,G.P.S.
TITLE Nucleotide sequence of four different isolates of porcine
circovirus detected in pigs with various clinical syndromes
JOURNAL Unpublished

```

```

REFERENCE 2 (bases 1 to 1768)
AUTHORS Hamel,A.L. and Nayar,G.P.S.
TITLE Direct Submission
JOURNAL Submitted (10-DEC-1998) Virology Laboratory, Manitoba Agriculture,
545 University Crescent, Winnipeg, Manitoba R3T 5S6, Canada
FEATURES
Location/Qualifiers
1..1768
/organism="porcine circovirus type 2-B"
/db_xref="taxon:85709"
/note="sequence obtained from several overlapping PCRs
using DNA extracted from lung, mesenteric lymph node and
tonsil of pig; similar to porcine circovirus sequence
presented in GenBank Accession Number AF027217; type-B
designation is based upon restriction endonuclease
digestion pattern"
repeat_region 13..35
/rpt_type=tandem
/rpt_unit=13..18
51..995
/note="ORF-1"
/codon_start=1
/product="putative Rep and coat protein"
/protein_id="AAD03086.1"
/db_xref="PID:g4106918"
/db_xref="GI:4106918"
/translation="MPSRNGRSGPQPHKRVFTLNPSDEKRIKELPISLFDYFI
VGEENEGRTPLHGFANFKQFENKVKWLGARCHEKAKGTDOONKEYCSKEGN
LLIEGAPRSQGRSLSSTAVSTLLESGSLVVAHQHPVTFVNRGLAELLKVSQGM
QRDKWNTNHHVIVGPPGCKSKAANFADPTTYWPKRNKWDGHEEVVVDIFY
GWLWDLLRLCDRPLTITGKGVFPFLARSILITSNQTPLEWYSSTAVPAVEALYR
RITSLVFWKNATEQSTEGGQFVTLSPPCPEPEYIN"
327..332
complement(357..671)
/codon_start=1
/product="ORF-3"
/protein_id="AAD03087.1"
/db_xref="PID:g4106919"
/db_xref="GI:4106919"
/translation="MVTIPLYSRWFPVCGFRVCKISSPFAFTTPKPHNDVDIGLPI
TLHFFPAHQFQSQPAEISDKRYRLLCGHQTALQGGTHSSRQVTPLSRGRSTF
NK"
complement(386..565)
/codon_start=1
/product="ORF-4"
/protein_id="AAD03088.1"
/db_xref="PID:g4106920"
/db_xref="GI:4106920"
/translation="MTWTLVFSQRCIFPLTFKSSASPRKFLTNVTGCCSATVTRLPL
SNKVLTAVDRLRCP"
553..732
/codon_start=1
/product="ORF-12"
/protein_id="AAD03089.1"
/db_xref="PID:g4106921"
/db_xref="GI:4106921"
/translation="MSTSLWGLGVVVKANGLLLIQTRKPHNGHLETSGGVTWVKW
LLLMFMACRGNLY"
complement(688..753)
/codon_start=1
/product="ORF-8"
/protein_id="AAD03094.1"
/db_xref="PID:g4106926"
/db_xref="GI:4106926"
/translation="MDIDHTVSVVDHPTAAASHKSHQ"
983..988
complement(989..1033)
/codon_start=1
/product="ORF-11"
/protein_id="AAD03090.1"
/db_xref="PID:g4106922"
/db_xref="GI:4106922"
/translation="MNNKNHYEVIKKTQ"
1016..1177
CDS

```



```

/codon_start=1
/product="ORF-5"
/protein_id="A403091.1"
/db_xref="PID:g4106923"
/db_xref="GI:4106923"
/translation="MVFIIHLGFKWGVKIKFSLEYIHGYTDIVVLVVFVTFERSAEA
YVHISRGL"
polyA_signal
complement(1022..1027)
complement(1256..1735)
/note="similar to Porcine circovirus ORF-2 encoded by the
sequence presented in GenBank Accession Number AF072717."
/codon_start=1
/product="ORF-2"
/protein_id="A403085.1"
/db_xref="PID:g4106917"
/db_xref="GI:4106917"
/translation="MTYPRRYYRRRRHPRSHLQILRRRLVHPHRRYVNRKNGI
FNRLSRFTGYTKATVTTPSWAVDMRFDVFPVGGTNRKISVFPFYIRKVK
VERWPCSPITQGRGVGSTAVILDNFVTKATALTYPVYNSRHTIPQPFVHSR"
complement(1522..1611)
/codon_start=1
/product="ORF-6"
/protein_id="A403084.1"
/db_xref="PID:g4106916"
/db_xref="GI:4106916"
/translation="MASSTPASPAPSDILSLPQSRPPGRWT"
1524..1631
/codon_start=1
/product="ORF-10"
/protein_id="A403092.1"
/db_xref="PID:g4106924"
/db_xref="GI:4106924"
/translation="MSTAQSGVTVVALTVPKVRRRLVLMKPFLLQR"
1682..1741
/codon_start=1
/product="ORF-7"
/protein_id="A403093.1"
/db_xref="PID:g4106925"
/db_xref="GI:4106925"
/translation="MAAGAVSSSAVTPPWIRHS"
join(1750..1768,1..13)
/note="putative replication site"
join(1762..1768,1..2)
/note="similar to the nonanucleotide motif of Porcine
circovirus presented in GenBank Accession Numbers AF072717
and U49186"
BASE COUNT 448 a 360 c 493 g 467 t
ORIGIN

alignment_scores:
  Quality: 531.00 Length: 104
  Ratio: 5.257 Gaps: 0
Percent Similarity: 97.115 Percent Identity: 94.231

alignment_block:
US-09-209-961-9 x AF112862/rev ..

Align seg 1/1 to reverse of: AF112862 from: 1 to: 1768

1 MetValThrIleProProLeuValPheArgTrpPheProValCysGlyPh 17
|||||
671 ATGGTAACCATCCACCACCTTTCTAGTGGTTTCCAGTATGTGTTT 622

17 eArgValCysLysIleSerProPheAlaPheThrProArgTrp 34
|||||
621 CCGGTCTGCAAAATAGCAGCCATTCTTTACCACCCAGGTGC 572

34 rHisAsnGluValTyrlleGlyPheProIleThrLeuLeuHisPhePro 50
|||||
571 CCCCAATGACGTGCATGTCTTCCATCAGCTTCTGCATTTCCC 522

51 AlaHisPheGlnLysPheSerGlnProAlaGluIlePheAspLysArgTy 67
|||||
521 GCTCACTTTCAAAAGTTTCAGCCAGCCGCGGAAATTTCTCACAAACGTTA 472

67 rArgValLeuLeuCysAsnGlyHisGlnAsnProAlaLeuGlnGlnGlyT 84
|||||
471 CAGGGTGTCTCTGCAACGGTCCACAGACTCCGCTCTCCAACAAGGTA 422

84 hHisSerSerArgGlnValThrProLeuSerLeuArgSerArgSerSer 100
|||||
421 CTCACAGCAGTAGACAGGTACCTCGTTGCTTGAGATCGAGGAGCTCC 372

101 ThrPheAsnLys 104
|||||
371 ACATTCAATAAG 360

seq_name: gb_v1:AF154679

seq_documentation_block:
LOCUS AF154679 1768 bp DNA circular VRL 13-JUN-1999
DEFINITION Porcine Circovirus, complete genome.
ACCESSION AF154679
VERSION 95052004
NID AF154679.1 GI:5052004
KEYWORDS
SOURCE
ORGANISM
porcine circovirus.
Viruses; ssDNA viruses; Circoviridae; Circovirus.
REFERENCE
1 (bases 1 to 1768)
AUTHORS
Kuo,T.Y., Chiou,Y.C. and Lai,S.S.
TITLE
Complete nucleotide sequences analysis of porcine circovirus
outbreak in Taiwan
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 1768)
AUTHORS
Kuo,T.Y., Chiou,Y.C. and Lai,S.S.
TITLE
Direct Submission
JOURNAL
Submitted (27-MAY-1999) Veterinary Medicine, National Taiwan
University, 142, Chousan Road, Taipei 106, Taiwan
FEATURES
Location/Qualifiers
source
1..1768
/organism="porcine circovirus"
/isolate="MLTW98"
/db_xref="taxon:46221"
/country="Taiwan"
51..995
/note="Putative Rep protein"
/codon_start=1
/product="ORF1"
/protein_id="AAD38398.1"
/db_xref="PID:95052005"
/db_xref="GI:5052005"
/translation="MPSKNGRSGPQPKRWVFTLNNPSEDERKKIRELPISLPCFI
VGEENEGRTPHLQGFANFYKOTFNKVKVFGARHIEAKGTDQONKEYCSKEGN
LLIEGAPRSQGRSDLSLSTVLSGLVTAQHPVTFVNFRLGLAELKVGKM
OKRDKTNHVIVGPGCGSKWAANFADPTTYWKPRNKWDCYHGEEVVIDFY
GWLFWDDLLRLCDRPLVETKGTGTPPLARSILITSNQTPLEWTSSTAVPAVEALYR
RITSLVFWKNATEQSTEGGFVTLSPPCPEFFYEIN"
BASE COUNT 451 a 367 c 495 g 455 t
ORIGIN

alignment_scores:
  Quality: 522.00 Length: 104
  Ratio: 5.168 Gaps: 0
Percent Similarity: 97.115 Percent Identity: 92.308

alignment_block:
US-09-209-961-9 x AF154679/rev ..

Align seg 1/1 to reverse of: AF154679 from: 1 to: 1768

1 MetValThrIleProProLeuValPheArgTrpPheProValCysGlyPh 17
|||||
671 ATGGTAACCATCCACCACCTTTCTAGTGGTTTCCAGTATGTGTTT 622

17 eArgValCysLysIleSerProPheAlaPheThrProArgTrp 34
|||||
621 CCGGTCTGCAAAATAGCAGCCATTCTTTACCACCCAGGTGC 572

34 rHisAsnGluValTyrlleGlyPheProIleThrLeuLeuHisPhePro 50
|||||
571 CCCCAATGACGTGCATGTCTTCCATCAGCTTCTGCATTTCCC 522

51 AlaHisPheGlnLysPheSerGlnProAlaGluIlePheAspLysArgTy 67
|||||
521 GCTCACTTTCAAAAGTTTCAGCCAGCCGCGGAAATTTCTCACAAACGTTA 472

67 rArgValLeuLeuCysAsnGlyHisGlnAsnProAlaLeuGlnGlnGlyT 84
|||||
471 CAGGGTGTCTCTGCAACGGTCCACAGACTCCGCTCTCCAACAAGGTA 422

84 hHisSerSerArgGlnValThrProLeuSerLeuArgSerArgSerSer 100
|||||
421 CTCACAGCAGTAGACAGGTACCTCGTTGCTTGAGATCGAGGAGCTCC 372

101 ThrPheAsnLys 104
|||||
371 ACATTCAATAAG 360

seq_name: gb_v1:AF154679

seq_documentation_block:
LOCUS AF154679 1768 bp DNA circular VRL 13-JUN-1999
DEFINITION Porcine Circovirus, complete genome.
ACCESSION AF154679
VERSION 95052004
NID AF154679.1 GI:5052004
KEYWORDS
SOURCE
ORGANISM
porcine circovirus.
Viruses; ssDNA viruses; Circoviridae; Circovirus.
REFERENCE
1 (bases 1 to 1768)
AUTHORS
Kuo,T.Y., Chiou,Y.C. and Lai,S.S.
TITLE
Complete nucleotide sequences analysis of porcine circovirus
outbreak in Taiwan
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 1768)
AUTHORS
Kuo,T.Y., Chiou,Y.C. and Lai,S.S.
TITLE
Direct Submission
JOURNAL
Submitted (27-MAY-1999) Veterinary Medicine, National Taiwan
University, 142, Chousan Road, Taipei 106, Taiwan
FEATURES
Location/Qualifiers
source
1..1768
/organism="porcine circovirus"
/isolate="MLTW98"
/db_xref="taxon:46221"
/country="Taiwan"
51..995
/note="Putative Rep protein"
/codon_start=1
/product="ORF1"
/protein_id="AAD38398.1"
/db_xref="PID:95052005"
/db_xref="GI:5052005"
/translation="MPSKNGRSGPQPKRWVFTLNNPSEDERKKIRELPISLPCFI
VGEENEGRTPHLQGFANFYKOTFNKVKVFGARHIEAKGTDQONKEYCSKEGN
LLIEGAPRSQGRSDLSLSTVLSGLVTAQHPVTFVNFRLGLAELKVGKM
OKRDKTNHVIVGPGCGSKWAANFADPTTYWKPRNKWDCYHGEEVVIDFY
GWLFWDDLLRLCDRPLVETKGTGTPPLARSILITSNQTPLEWTSSTAVPAVEALYR
RITSLVFWKNATEQSTEGGFVTLSPPCPEFFYEIN"
BASE COUNT 451 a 367 c 495 g 455 t
ORIGIN

```

```

17 eArgValCysLysLeSerSerPropheAlaPheThrThrProArgTrpP 34
621 CCGGGTCTGCAAAATTAGCAGCCATTGCTTTTACCACACCCAGGTGGC 572
34 roHisAsnGluValTyIleGlyPheProIleThrLeuLeuHisPhePro 50
571 CCCAATGACGTGTACATTGCTTCCAATCAGCTTCTGCATTTTCCC 522
51 AlaHisPheGlnLysPheSerGlnProAlaGluIlePheAspLysArgTy 67
521 GCTCATTCTCAAAAGTTAGCAGCCCGCGGAAATTTCTGACAAACGTTA 472
67 rArgValLeuLeuCysAsnGlyHisGlnAsnProAlaLeuGlnGlyT 84
471 CAGGGTCTGCTGTGCAACGGTACACGACTCCGCTCTCCACCAAGGTA 422
84 hrHisSerSerArgGlnValThrProLeuSerLeuArgSerArgSerSer 100
421 CTCACAGCAGTAGACAGGTCACTGCGTTGTCTTCCAAATCAGCTTCTGCATTTTCCC 522
101 ThrPheAsnLys 104
371 ACATCGATAAG 360

seq_name: gb_vi:AF166528
seq_documentation_block:
LOCUS AF166528 1768 bp mRNA VRL 19-AUG-1999
DEFINITION Porcine circovirus complete genome.
ACCESSION AF166528
NID 95739338
VERSION AF166528.1 GI:5739338
KEYWORDS
SOURCE
ORGANISM
porcine circovirus.
porcine circovirus
Viruses; ssDNA viruses; Circoviridae; Circovirus.
REFERENCE 1 (bases 1 to 1768)
AUTHORS Yang, K.H., Lee, Y.F., Chao, D.S., Shieh, Y.C. and Lai, S.S.
TITLE Complete nucleotide sequences of porcine circovirus Tainan strand outbreak in Taiwan
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1768)
AUTHORS Yang, K.H., Lee, Y.F., Chao, D.S., Shieh, Y.C. and Lai, S.S.
TITLE Direct Submission
JOURNAL Submitted (08-JUL-1999) Veterinary Medicine, National Chia-Yi Institute of Technology, 300 Shei Fu Road, Lu Liao Li, Chia-Yi City 600, Taiwan
FEATURES
Location/Qualifiers
1..1768
/organism="porcine circovirus"
/virusion
/strain="Tainan"
/db_xref="taxon:46221"
/country="Taiwan"
51..995
/codon_start=1
/product="putative Rep protein"
/protein_id="AAD50432.1"
/db_xref="PID:5739339"
/db_xref="GI:5739339"
/translation="MPSKNGSGPQPKRWVFTLNNPSEDERKKIRELPISLFDYFI
VGEENGEGRTHLQGFANFYKQTFNKVYFGRCHIEKAGKGDQDNKEYCSREGN
LLIECGAPRSQGRSDLSLSTLLESGLVTVAEQHPVTFVNFRLAELLKVSQGM
QKRDKNVHVTPGCGSKAAANFADPTTYKPRNKWDYHGVEVYIDDFY
GWLWDLLRLCDRYPLVETRGVTPFLARSLITSNOTPLEWYSSTAPPAVEALYR
RITSLVFNKNAEQSTEEGGQFVTLSPCPPEFYINY"
BASE COUNT 453 a 367 c 492 g 456 t
ORIGIN

```

alignment\_scores:

Quality: 522.00

Length: 104



\*\*\*\*\*  
W P S R E H  
\*\*\*\*\* (TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Wed Dec 22 10:25:26 1999; MasPar time 12.23 Seconds  
102.575 Million cell updates/sec  
Tabular output not generated.

Title: >US-09-209-961-20  
Description: (1-59) from US09209961.pep  
Perfect Score: 471  
Sequence: 1 MYTSLWGLGVKRWANGLLIL.....MVKKWLMTFMAGCRGMIV 59

Scoring table: PAM 150  
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq35  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39

Statistics: Mean 26.466; Variance 97.683; scale 0.271

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	78	16.6	365	13 R69518	Prostaglandin-EP3-9 r	2.36e+01
2	78	16.6	380	13 R69517	Prostaglandin-EP3-21	2.36e+01
3	78	16.6	398	13 R69516	Prostaglandin-EP3-1p	2.36e+01
4	78	16.6	393	32 W57411	Human prostaglandin E	2.36e+01
5	78	16.6	402	32 W57410	Human EP3-V receptor	2.36e+01
6	78	16.6	427	6 R32085	Benzene dioxygenase p	2.36e+01
7	78	16.6	441	32 W28236	Amino acid sequence o	2.36e+01
8	78	16.6	672	3 P60219	Sequence of thermosta	2.36e+01
9	78	16.6	672	2 P71200	Thermostable beta-gal	2.36e+01
10	77	16.3	534	2 P70324	Sequence of Ricin com	2.87e+01
11	77	16.3	4630	23 W19629	Streptomyces venezuel	2.87e+01
12	76	16.1	540	24 W25143	Castor oil plant aggl	3.49e+01
13	76	16.1	540	23 W21706	R. communis agglutini	3.49e+01
14	75	15.9	189	34 W29746	TNF related endotheli	4.24e+01
15	75	15.9	305	29 W55451	H. pylori ORF 02ae116	4.24e+01
16	75	15.9	332	29 W55210	H. pylori ORF 02ae116	4.24e+01

17 74 15.7 231 3 260030 Sequence of ricin B c  
18 74 15.7 231 1 R31329 The ricin B chain fro  
19 74 15.7 251 1 P31887 Amino acid sequence o  
20 74 15.7 251 1 P90837 Sequence encoded by i  
21 74 15.7 262 1 P81852 Ricin B mutein with d  
22 74 15.7 262 1 P81849 Ricin B mutein with d  
23 74 15.7 262 1 P90020 B chain of ricin D  
24 74 15.7 262 1 P82853 Ricin B mutein with d  
25 74 15.7 262 1 P82896 Ricin B mutein with d  
26 74 15.7 262 1 P82865 Generic sequence for  
27 74 15.7 262 1 P82851 Ricin B mutein with d  
28 74 15.7 262 1 P82850 Ricin B mutein with d  
29 74 15.7 306 35 W71512 Helicobacter polypt  
30 74 15.7 317 3 P60507 Human apolipoprotein  
31 74 15.7 317 16 R86791 Human apolipoprotein  
32 74 15.7 317 3 P51204 Human apolipoprotein E  
33 74 15.7 562 1 P90079 Ricin D  
34 74 15.7 565 3 P50166 Sequence of preproric  
35 74 15.7 565 3 P60240 Preproricin  
36 74 15.7 574 2 P70325 Sequence of Ricinus c  
37 74 15.7 574 1 P94793 DNA sequence of ricin  
38 74 15.7 576 2 P70326 Sequence of Ricinus c  
39 74 15.7 576 26 W25787 Castorbean ricin  
40 73 15.5 73 3 R12605 SIB 136 intestinal mu  
41 73 15.5 102 14 W74204 Human death associate  
42 73 15.5 102 36 W71366 Death associated prot  
43 72 15.3 236 26 P60039 Sequence of ricin B c  
44 72 15.3 236 2 P91888 Sequence of ricin B c  
45 72 15.3 584 2 R06520 Microspore-specific c

ALIGNMENTS

RESULT 1  
ID R69518 standard; Protein: 365 AA.  
AC R69518;  
DE 02-SEP-1994 (first entry)  
DE Prostaglandin-EP3-9 receptor;  
AW human uterus prostaglandin-EP3-9 receptor;  
KW prostaglandin-EP3 receptor-agonist;  
KW prostaglandin-EP3 receptor-antagonist;  
CS Homo sapiens.  
FH Key Location/Qualifiers  
FT modified\_site 18  
FT modified\_site 36 /note= "N-glycosylation site"  
FT modified\_site 217 /note= "N-glycosylation site"  
FT modified\_site 308 /note= "N-glycosylation site"  
FT modified\_site 338..345 /note= "N-glycosylation site"  
FT region /note= "conserved sequence in EP3 receptors"  
FT WO95JJ552-A.  
FT 05-JAN-1995.  
FT 09-JUN-1994; CAC020.  
FT 25-JUN-1993; US-083746.  
FT (MERK) MERCK PROSST CANADA INC.  
FT Abramowitz M, Adam M, Boile Y, Metters K, Rushmore TH;  
FT WPI: 95-052012/07.  
FT N-PDB; Q81951.  
FT New prostaglandin EP3 receptors and DNA - used partic. to  
FT identify modulators of prostaglandin receptor activity for  
FT treatment of diseases.  
FT Claim 2; Page 43-75; 64pp; English.  
FT The sequence represents a human uterus prostaglandin-EP3-9  
FT receptor (mol.wt. 40,507). The sequence contains 4 conserved  
FT N-glycosylation sites in putative extracellular regions, and a  
FT conserved region (R69515) in transmembrane region-VII common to  
FT other EP3 receptors. Conserved cys residues are found in  
FT exofacial loops 1 and 2, and a conserved Arg found in all  
FT eicosanoid receptors is found in transmembrane region-VII. The  
FT receptor may be used to identify receptor-agonists and

RESULT 3  
ID R69516 standard; Protein; 390 AA.  
AC R69516;  
02-SEP-1995 (first entry)  
Prostaglandin-EP3-alpha receptor.

DR New human prostaglandin EP3 receptor(s) - useful for treatment and  
PT prevention of, e.g. inflammation  
PS Claim 8; Pages 20-21; 27pp; Japanese.  
CC This represents a human EP3-VI receptor. A replication or expression  
CC vector comprising cDNA sequences encoding EP-3V or EP3-3VI can be  
CC used to transform a host cell. The host cell is cultured and the polypep-  
CC tide can be recovered from the culture medium. The polypeptides combine  
CC specifically with a prostaglandin  $g_{p2}$  receptor and can be used as  
CC agonists.

CC preventive and treating agent for inflammation.  
SQ Sequence 393 AA;

Query Match 16.6%; Score 78; DB 32; Length 393;  
Best Local Similarity 31.3%; Pred. No. 2.36e+01;  
Matches 10; Conservative 9; Mismatches 12; Indels 1; Gaps 1;

Db 157 lairaphyashmktravllgwlavlaf 188  
QY 20 LQTRKPH-TGNHLKTSGGMTVMVKWLLMTF 50

## RESULT 5

ID W57410 standard; Protein; 402 AA.  
AC W57410;  
DT 19-AUG-1998 (first entry)  
DE Human EP3-V receptor.  
KW Prostaglandin E2 receptor; EP3-V receptor; human; treatment;  
OS Homo sapiens.  
PN J10113185-A.  
PD 06-MAY-1998.  
PF 14-OCT-1996; JP-291150.  
PR 14-OCT-1996; JP-291150.  
PA (ONOX) ONO PHARM CO LTD.  
DR WPI; 98-315474/28.  
DR N-PSDB; V29610.  
PT New human prostaglandin EP3 receptor(s) - useful for treatment and prevention of, e.g. inflammation.  
PS Claim 2; Pages 16-17; 27pp; Japanese.  
CC This represents a human EP3-V receptor. A replication or expression vector comprising cDNA sequences encoding EP3-V or EP3-3VI can be used to transform a host cell. The host cell is cultured and the polypeptides can be recovered from the culture medium. The polypeptides combine specifically with a prostaglandin PGE2 receptor and can be used as a preventive and treating agent for inflammation.  
SQ Sequence 402 AA;

Query Match 16.6%; Score 78; DB 32; Length 402;  
Best Local Similarity 31.3%; Pred. No. 2.36e+01;  
Matches 10; Conservative 9; Mismatches 12; Indels 1; Gaps 1;

Db 157 lairaphyashmktravllgwlavlaf 188  
QY 20 LQTRKPH-TGNHLKTSGGMTVMVKWLLMTF 50

## RESULT 6

ID R32085 standard; Protein; 427 AA.  
AC R32085;  
DT 07-JUN-1993 (first entry)  
DE Benzene dioxygenase polypeptide I.  
KW Benzene; dioxygenase; cysbenzene glycol; transformation; BDO.  
OS Pseudomonas aeruginosa.  
PN J05003785-A.  
PD 14-JAN-1993.  
PF 30-OCT-1991; 284795.  
PR 16-APR-1991; JP-084170.  
PA (SAOC) MERCIAN CORP.  
DR WPI; 93-054795/07.  
DR N-PSDB; Q36395.  
PT Microbe DNA - contg. benzene di-oxygenase gene, cis:benzene glycol dehydrogenase gene and catechol 2,3-oxygenase gene  
PS Claim 12; Page 4-5; 20pp; Japanese.  
CC The sequences given in R32085-89 represent five regions of a benzene dioxygenase (BDO) polypeptide. Cysbenzene glycol can be produced CC effectively from benzene using transformants prepared using the DNA CC encoding these fragments.  
SQ Sequence 427 AA;

Query Match 16.6%; Score 78; DB 6; Length 427;  
Best Local Similarity 38.6%; Pred. No. 2.36e+01;  
Matches 17; Conservative 8; Mismatches 16; Indels 3; Gaps 3;

Db 60 lgheshypetgdfatymgedpvmvirkdksikvflnqcrqm 103  
QY 17 LLILQTRKPH-TGNHLKT-SG-GMVTMVK-KWLLMTFMAGCRGM 57

## RESULT 7

ID W28236 standard; Protein; 441 AA.  
AC W28236;  
DT 07-SEP-1998 (first entry)  
DE Amino acid sequence of a mercuric reductase.  
KW Staphylococcus aureus protein; ribozyme; antisense sequence; control;  
KW Staphylococcal gene; regulatory element; bacterial gene expression;  
KW vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;  
KW toxic shock syndrome; mercuric reductase.  
OS Staphylococcus aureus.  
FH Key Location/Qualifiers  
FT Misc difference: 16 /note= "Residue not specified"  
PN WO9730070-A1  
PD 21-AUG-1997.  
PF 19-FEB-1997; U02318.  
PR 20-FEB-1996; US-011888.  
PS (SMIK) SMITHKLINE BEECHAM CORP.  
PI Black Mt, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO,  
PI Pratt JM, Reichard RW, Rosenberg M, Ward JM;  
DR WPI; 97-124969/39.  
DR N-PSDB; T84151.  
PT Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used to isolate antimicrobial compounds, and in vaccines against S. aureus infection  
PS Claim 1; Pages 547-548; 939pp; English.  
CC The present sequence represents a Staphylococcus aureus protein, that, based on homology with another Staphylococcus aureus protein, is believed to be a mercuric reductase. The DNA sequence was isolated from a library of clones of S. aureus WCUH 29 in Escherichia coli. The DNA sequence can be used in the construction of Staphylococcal genes. The DNA sequence is also useful as a source of regulatory elements for the control of bacterial gene expression. The present protein may be used to produce vaccines to enable a host to produce specific antibodies with antibacterial action. These vaccines and antibodies would protect a host against invasion by S. aureus, and conditions relating to Staphylococcal infection, e.g. Staphylococcal food poisoning, scaled skin syndrome, and toxic shock syndrome.  
SQ Sequence 441 AA;

Query Match 16.6%; Score 78; DB 32; Length 441;  
Best Local Similarity 26.7%; Pred. No. 2.36e+01;  
Matches 16; Conservative 14; Mismatches 28; Indels 2; Gaps 2;

Db 294 lyievngqevleadqvlvratgrkptet-lnlessgvtgkkgvltneylgtsmriy 352  
QY 1 MYTSLWGLGVYKANGLLILOTRKPH-TGNHLKTSGGMTVMVKWLLMT-FMAGCRGM 59

## RESULT 8

ID P60219 standard; Protein; 672 AA.  
AC P60219;  
DT 24-AUG-1991 (first entry)  
DE Sequence of thermostable beta-galactosidase.  
KW Low lactose milk; galactose; lactose intolerance; food processing.  
OS Bacillus sphaerothermophilus.  
PN EP-176971-A.  
PD 09-APR-1986.  
PF 29-SEP-1984 JP-202965.  
PR 24-MAY-1986; JP-110565.  
PA (WAKA) WAKAMOTO PHARM KK.  
DR Hirata H, Ikada H, Negoro S;  
DR WPI; 86-G53363/15.  
DR N-PSDB; N60150.  
PT Thermostable beta-galactosidase from bacillus subtilis



PI Hey TD, Morgan-AER. Walsh TA; -----  
DR WPI: 97-309831/28.  
PT Inactive precursor of maize ribosome-inactivating protein - also  
PT chimeric r.bosomer-inactivating protein precursors containing  
PT internal linker sequences

PS Claim 2; Column 119-124; 12lpp; English.

CC The sequences given in W21698-710 represent Ribosome Inactivating  
CC Proteins (RIP's); which may be used in the construction of the  
CC proRIP of the invention. The proRIP has a selectively removable,  
CC internal peptide linker. The precursor sequence is incapable of  
CC inactivating eukaryotic ribosomes, but can be converted by removal  
CC of the linker into a protein having alpha and beta fragments and being  
CC capable of inactivating eukaryotic ribosomes. RIPs are potent  
CC inhibitors of eukaryotic protein synthesis. They possess a highly  
CC specific N-glycosidase activity which cleaves the glycosidic bond of  
CC adenine 4324 of rat liver ribosomal 28S RNA. RIP's selectively inhibit  
CC cellular proliferation of cells, e.g. cancer cells and HIV-infected T  
CC cells. The inactive proRIP proteins make it possible to provide protein  
CC synthesis inhibitors with uses in practical and improved ways not before  
CC possible. The RIP can be used to make cytotoxic conjugates.  
SQ Sequence 540 AA;

Query Match 16.1%; Score 76; DB 23; Length 540;  
Best Local Similarity 17.1%; Pred. NO. 3.49e+01;  
Matches 7; Conservative 14; Mismatches 20; Indels 0; Gaps 0;

Db 370 iwdnrtlnprsgivlaatsngstklvtqtniyavsqgw 110  
QY :|:::|::|::|::|::|::|::|::|::|::|::|:  
5 LMGHLGVVKAGLLILOTKPKHTNHLTSGGMVTMKWL 45

RESULT 14

ID W29746 standard; Protein; 189 AA.

AC W29746;  
DT DT 27-OCT-1998 (first entry)

DF TNF related endothelium proliferative agent protein 2.  
KW Endothelium proliferative agent; TREPA; wound healing; cancer;  
KW tissue grafting; vascularisation; apoptosis; autoimmune; birth control.  
OS Homo sapiens.  
PN WO9835061-A2.  
PD 13-AUG-1998.  
PF 12-FEB-1998; U02859.  
PR 10-FEB-1998; US-021706.  
PR 12-FEB-1997; US-798692.  
PA (ABBO) ABBOTT LAB.  
PI Wiley SR;  
DR WPI: 98-447255/38.  
PT Detecting nucleic acid encoding TREPA - useful for diagnosis and  
PT treatment of autoimmune disease, tumours and inflammation

PS Claim 16; Page 125-6; 142pp; English.

CC The TNF-related endothelium proliferative agent (TREPA), or its  
CC activators or agonists, are used to treat a deficit of TREPA, e.g. to  
CC promote wound-healing or tissue grafting, by promoting vascularisation,  
CC also to induce apoptosis for treating cancer and eliminating autoreactive  
CC T cells, as an adjunct to cancer chemotherapy or antiviral treatment.  
CC TREPA peptides can also be used to target cytotoxic agents or for  
CC affinity isolation of the corresponding receptor, the nucleic acid for  
CC which can be used to transform tumour cells to render them more  
CC responsive to TREPA and to screen for TREPA mimics.  
CC Ribozymes, antisense RNA, antibodies or peptides, are used to treat  
CC TREPA-associated diseases, e.g. tumours and metastases (by inhibiting  
CC vascularisation), inflammation or a wide range of autoimmune conditions,  
CC conditions involving abnormal stimulation of epithelial cells (e.g.  
CC atherosclerosis), for birth control (inhibiting ovulation and placental  
CC formation) or other angiogenic conditions (e.g. ulcers).  
SQ Sequence 189 AA;

Query Match 15.9%; Score 75; DB 34; Length 189;  
Best Local Similarity 45.0%; Pred. NO. 4.24e+01;  
Matches 9; Conservative 5; Mismatches 4; Indels 2; Gaps 2;

Db 1 mvmirtwrllpmvllaayc 20  
II:::|::|::|::|::|::|::|::|::|::|::|:



QY 37 MVTMVKKW-LLMTFMAG-C 54

## RESULT 15

ID W55451 standard; Protein; 305 AA.  
 AC W55451;  
 DT 24-JUN-1998 (first entry)  
 DE H. pylori ORF 02aell1612\_23598175\_f1\_2 secreted protein.  
 KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
 KW identification; binding compound; bacteria; life cycle; activator;  
 KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.  
 OS Helicobacter pylori.  
 PN WO9737044-A1.  
 PD 09-OCT-1997.  
 PF 27-MAR-1997; U052223.  
 PR 06-DEC-1996; US-761318.  
 PR 29-MAR-1996; US-625811.  
 PR 02-APR-1996; US-758731.  
 PR 25-OCT-1996; US-736905.  
 PR 28-OCT-1996; US-738859.  
 PA (ASTR ) ASTRA AB.  
 PI Alm RA, Smith D;  
 DR WPI; 97-503122/46.  
 DR N-PSDB; V24860.  
 PT Helicobacter pylori nucleic acid sequences and encoded  
 PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori  
 PT infection and for diagnosis of H. pylori infection  
 PS Claims 14, 94; Page 658; 1145pp; English.  
 CC This sequence is a H. pylori secreted protein.  
 CC The protein may be used in a vaccine to prevent or treat H. pylori  
 CC infection or to identify H. pylori polypeptide binding compounds. The  
 CC useful as potential H. pylori life cycle activators or inhibitors. The  
 CC DNA and probes derived from it may be used for the identification of  
 CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic  
 CC acid sequences complementary to the DNA act as antisense sequences and  
 CC can be used to prevent the translation of H. pylori mRNA. Antibodies  
 CC against the protein can be used in immunoassays to evaluate the abundance  
 CC and distribution of H. pylori-specific antigens. The genomic sequence of  
 CC H. pylori (ATCC 53679) was determined from overlapping contigs generated  
 CC by mechanically shearing the bacterial DNA. The sequences were analysed  
 CC for ORF of at least 180 nucleotides, and the predicted coding regions  
 CC defined by computer evaluation. To identify likely H. pylori antigens for  
 CC vaccine development, the amino acid sequences predicted from various ORF  
 CC were analysed for significant homology to other known or exported  
 CC membrane proteins. Having identified and determined the sequences of  
 CC interest, particular regions can be isolated from H. pylori by PCR  
 CC amplification for recombinant polypeptide production, e.g. in E. coli  
 CC hosts.  
 CC Sequence 305 AA;

Query Match 15.9%; Score 75; DB 29; Length 305;  
 Best Local Similarity 42.9%; Pred. No. 4.24e+01;  
 Matches 9; Conservative 6; Mismatches 5; Indels 1; Gaps 1;

Db 1 mikswtkkwliflmscfig 21  
 | : |||: :||| :|  
 QY 37 MVTM-VKKWLLMTFMAGCRG 56

Search completed: Wed Dec 22 10:25:43 1999  
 Job time : 17 secs.

\*\*\*\*\*  
W O S R E L H  
\*\*\*\*\*  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (C) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Dec 22 10:27:23 1999; MasPar time 3.60 Seconds  
Tabular output not generated. 195.463 Million cell updates/set

Title: >US-09-209-961-20  
Description: (1-59) from US09209961.pep  
Perfect Score: 471  
Sequence: 1 MYTSLGHLGVKRWANGLLIL.....MVKKLLMTFMAGCRGMIV 59

Scoring table: PAM 150  
Gap 11

Searched: 122461 seqs, 11912985 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-issued  
1:5A\_COMB 2:5B\_COMB 3:PC9\_COMB 4:backfiles1

Statistics: Mean 24.934; Variance 98.568; scale 0.253

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result		Query		Match		Length		DB		ID		Description		Pred. No.	
1	76	16.1	540	1	US-08-485-	Sequence 77,	Applicati	2.35e+01							
2	76	16.1	540	1	US-08-378-	Sequence 77,	Applicati	2.35e+01							
3	74	15.7	220	1	US-08-726-	Sequence 29,	Applicati	3.41e+01							
4	74	15.7	317	1	US-07-709-	Sequence 2,	Applicati	3.41e+01							
5	72	15.3	414	1	US-08-002-	Sequence 19,	Applicati	4.93e+01							
6	72	15.3	414	1	US-08-002-	Sequence 17,	Applicati	4.93e+01							
7	72	15.3	414	1	US-08-002-	Sequence 13,	Applicati	4.93e+01							
8	72	15.3	484	1	US-08-030-	Sequence 8,	Applicati	4.93e+01							
9	71	15.1	529	1	US-08-496-	Sequence 2,	Applicati	5.93e+01							
10	70	14.9	510	1	US-08-278-	Sequence 4,	Applicati	7.12e+01							
11	69	14.6	361	2	US-08-685-	Sequence 4,	Applicati	8.54e+01							
12	69	14.6	361	1	US-08-390-	Sequence 4,	Applicati	8.54e+01							
13	69	14.6	365	1	US-08-390-	Sequence 2,	Applicati	8.54e+01							
14	69	14.6	365	2	US-08-685-	Sequence 2,	Applicati	8.54e+01							
15	69	14.6	532	3	PC9-US95-0	Sequence 6,	Applicati	8.54e+01							
16	69	14.6	532	1	US-08-481-	Sequence 6,	Applicati	8.54e+01							
17	68	14.4	235	1	US-07-932-	Sequence 1,	Applicati	1.02e+02							
18	68	14.4	528	2	US-08-466-	Sequence 2,	Applicati	1.02e+02							
19	68	14.4	528	2	US-08-700-	Sequence 2,	Applicati	1.02e+02							
20	67	14.2	969	2	US-08-284-	Sequence 2,	Applicati	1.22e+02							
21	67	14.2	969	3	PC9-US93-0	Sequence 2,	Applicati	1.22e+02							
22	67	14.2	3200	2	US-08-477-	Sequence 8,	Applicati	1.22e+02							
23	66	14.0	181	2	US-08-353-	Sequence 74,	Applicati	1.46e+02							

24 66 14.0 212 2 US-08-663- Sequence 2, Applicatio 1.46e+02  
25 66 14.0 216 2 US-08-713- Sequence 7, Applicatio 1.46e+02  
26 66 14.0 292 2 US-08-353- Sequence 11, Applicati 1.46e+02  
27 66 14.0 349 1 US-08-459- Sequence 33, Applicati 1.46e+02  
28 66 14.0 349 1 US-08-197- Sequence 33, Applicati 1.46e+02  
29 66 14.0 349 1 US-08-459- Sequence 33, Applicati 1.46e+02  
30 66 14.0 353 1 US-08-137- Sequence 43, Applicati 1.46e+02  
31 66 14.0 353 1 US-08-459- Sequence 43, Applicati 1.46e+02  
32 66 14.0 353 1 US-08-459- Sequence 43, Applicati 1.46e+02  
33 66 14.0 894 1 US-08-486- Sequence 4, Applicatio 1.46e+02  
34 66 14.0 894 1 US-08-117- Sequence 4, Applicatio 1.46e+02  
35 65 13.8 251 2 US-08-602- Sequence 36, Applicati 1.75e+02  
36 65 13.8 732 3 PC9-US93-1 Sequence 2, Applicatio 1.75e+02  
37 65 13.8 884 2 US-08-982- Sequence 2, Applicatio 1.75e+02  
38 65 13.8 884 2 US-08-465- Sequence 9, Applicatio 1.75e+02  
39 64 13.6 42 2 US-08-472- Sequence 9, Applicatio 2.09e+02  
40 64 13.6 291 1 US-08-467- Sequence 11, Applicati 2.09e+02  
41 64 13.6 291 3 PC9-US96-0 Sequence 11, Applicati 2.09e+02  
42 64 13.6 1408 2 US-08-612- Sequence 2, Applicatio 2.09e+02  
43 64 13.6 1565 4 5352450-2 Patent No. 5352450.  
44 64 13.6 2183 1 US-08-348- Sequence 7, Applicatio 2.09e+02  
45 64 13.6 3077 4 5223423-2 Patent No. 5223423.

ALIGNMENTS

RESULT 1  
ID US-08-485-286-77 STANDARD; PRT; 540 AA.  
XX AC XXXXXX  
YA  
YT

Sequence 77, Application US/08485286

Sequence 77, Application US/08485286  
Patent No. 5646026  
Patent No. 5646026 5646119

GENERAL INFORMATION:  
APPLICANT: WALSH, TERENCE A  
APPLICANT: HEY, TIMOTHY D  
APPLICANT: MORGAN, ALICE ER  
TITLE OF INVENTION: RIBOSOME-INACTIVATING PROTEINS, INACTIVE  
TITLE OF INVENTION: PRECURSOR FORMS THEREOF, A PROCESS FOR MAKING A METHOD  
TITLE OF INVENTION: USING  
NUMBER OF SEQUENCES: 81

CORRESPONDENCE ADDRESS:  
ADDRESSEE: ANDREA T. BORUCKI  
STREET: 19330 ZIONSVILLE ROAD  
CITY: INDIANAPOLIS  
STATE: IN  
COUNTRY: US  
ZIP: 46268  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,286  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/378761  
FILING DATE: 26-JAN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: BORUCKI, ANDREA T  
REGISTRATION NUMBER: 33651  
REFERENCE/DOCKET NUMBER: 38272B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (317) 337-4846  
INFORMATION FOR SEQ ID NO: 77:  
SEQUENCE CHARACTERISTICS:

CC LENGTH: 540 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
SQ SEQUENCE 540 AA; 60168 MW; 1511677 CN;

Query Match 16.1%; Score 76; DB 1; Length 540;  
Best Local Similarity 17.1%; Pred. No. 2.35e+01;  
Matches 7; Conservative 14; Mismatches 20; Indels 0; Gaps 0;

Db 370 IWDNRTIINPRSGVLATSGNSGKLVQTNIVAVSOGWL 410  
Qy 5 LWGLGVVKGKANGLLILQPKPHGTGNHLATSGGWTWVKWL 45

RESULT 2

ID US-08-378-761A-77 STANDARD; PRT; 540 AA.  
XX AC xxxxxx

Sequence 77, Application US/08378761A

Sequence 77, Application US/08378761A

Patent No. 5635384

GENERAL INFORMATION:

APPLICANT: WALSH, TERENCE A

APPLICANT: HEY, TIMOTHY D

APPLICANT: MORGAN, ALICE ER

TITLE OF INVENTION: RIBOSOME-INACTIVATING PROTEINS, INACTIVE

TITLE OF INVENTION: PRECURSOR FORMS THEREOF, A PROCESS FOR MAKING A METHOD OF

TITLE OF INVENTION: USING

NUMBER OF SEQUENCES: 81

CORRESPONDENCE ADDRESS:

ADDRESSEE: ANDREA T. BORUCKI

STREET: 9330 ZIONSVILLE ROAD

CITY: INDIANAPOLIS

STATE: IN

COUNTRY: US

ZIP: 46268

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/378,761A

FILING DATE: 26-JAN-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: BORUCKI, ANDREA T

REGISTRATION NUMBER: 33651

REFERENCE/DOCKET NUMBER: 38272B

TELEPHONE: (317) 337-4846

INFORMATION FOR SEQ ID NO: 77:

SEQUENCE CHARACTERISTICS:

LENGTH: 540 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE 540 AA; 60168 MW; 1511677 CN;

Query Match 16.1%; Score 76; DB 1; Length 540;  
Best Local Similarity 17.1%; Pred. No. 2.35e+01;  
Matches 7; Conservative 14; Mismatches 20; Indels 0; Gaps 0;

Db 370 IWDNRTIINPRSGVLATSGNSGKLVQTNIVAVSOGWL 410  
Qy 5 LWGLGVVKGKANGLLILQPKPHGTGNHLATSGGWTWVKWL 45

RESULT 3

ID US-08-726-306A-29 STANDARD; PRT; 220 AA.

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX Sequence 2, Application US/07709949  
CC Patent No. 5472858  
CC GENERAL INFORMATION:  
CC APPLICANT: Attie, Alan D  
CC APPLICANT: Gretch, Daniel G  
CC APPLICANT: Sturley, Stephen L  
CC APPLICANT: Beckage, Nancy E  
CC TITLE OF INVENTION: Production of Recombinant Proteins in  
CC TITLE OF INVENTION: Insect Larvae  
CC NUMBER OF SEQUENCES: 2  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Quarles & Brady  
CC STREET: P.O. Box 2113  
CC CITY: Madison  
CC STATE: WI  
CC COUNTRY: USA  
CC ZIP: 53701-2113  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/07709,949  
CC FILING DATE: 19910604  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Seay, Nicholas J  
CC REGISTRATION NUMBER: 27,386  
CC REFERENCE/DOCKET NUMBER: 9629691801  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (608) 251-5000  
CC TELEFAX: (608) 251-9166  
CC INFORMATION FOR SEQ ID NO: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 317 amino acids  
CC TYPE: AMINO ACID  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 317 AA; 36210 MW; 438748 CN;  
  
Query Match 15.7%; Score 74; DB 1; Length 317;  
Best Local Similarity 50.0%; Pred. No. 3.41e+01;  
Matches 8; Conservative 6; Mismatches 1; Indels 1; Gaps 1;  
  
Db 5 WAALLVTELAGCAKV 20  
QY 44 WL-LLTMTFAGCRGMI 58  
  
RESULT 5  
ID US-08-002-202-19 STANDARD; PRT: 414 AA.  
XX  
AC xxxxxx  
XX  
XX  
DT  
XX  
XX  
DE Sequence 19, Application US/08002202  
XX  
XX Sequence 19, Application US/08002202  
CC Patent No. 5604201  
CC GENERAL INFORMATION:  
CC APPLICANT: Thomas, Garry  
CC APPLICANT: Anderson, Eric D  
CC APPLICANT: Thomas, Laurel  
CC APPLICANT: Hayflick, Joel S  
CC TITLE OF INVENTION: Methods and Reagents for Inhibiting  
CC TITLE OF INVENTION: Furin Endoprotease  
CC NUMBER OF SEQUENCES: 21  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Allegretti and Witcoff, Ltd.  
CC STREET: 10 South Wacker Drive, Suite 3000  
CC CITY: Chicago  
CC STATE: Illinois  
CC COUNTRY: USA  
CC ZIP: 60606  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/002,202  
CC FILING DATE: 08-JAN-1993  
CC CLASSIFICATION: 514

CC CITY: Chicago  
CC STATE: Illinois  
CC COUNTRY: USA  
CC ZIP: 60606  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/002,202  
CC FILING DATE: 08-JAN-1993  
CC CLASSIFICATION: 514  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: No. 5604201nan, Kevin E  
CC REGISTRATION NUMBER: 35,30003  
CC REFERENCE/DOCKET NUMBER: 92,448  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 312-715-1000  
CC TELEFAX: 312-715-1234  
CC TELEFAX: 910-221-5317  
CC INFORMATION FOR SEQ ID NO: 19:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 414 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 414 AA; 46463 MW; 926194 CN;  
  
Query Match 15.3%; Score 72; DB 1; Length 414;  
Best Local Similarity 31.3%; Pred. No. 4.93e+01;  
Matches 5; Conservative 7; Mismatches 4; Indels 0; Gaps 0;  
  
Db 258 WVLLMKYLGNTAMFF 273  
QY 44 WLLTMTFAGCRGMI 59  
  
RESULT 6  
ID US-08-002-202-17 STANDARD; PRT: 414 AA.  
XX  
XX xxxxxx  
XX  
XX  
XX  
DE Sequence 17, Application US/08002202  
XX  
XX Sequence 17, Application US/08002202  
CC Patent No. 5604201  
CC GENERAL INFORMATION:  
CC APPLICANT: Thomas, Garry  
CC APPLICANT: Anderson, Eric D  
CC APPLICANT: Thomas, Laurel  
CC APPLICANT: Hayflick, Joel S  
CC TITLE OF INVENTION: Methods and Reagents for Inhibiting  
CC TITLE OF INVENTION: Furin Endoprotease  
CC NUMBER OF SEQUENCES: 21  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Allegretti and Witcoff, Ltd.  
CC STREET: 10 South Wacker Drive, Suite 3000  
CC CITY: Chicago  
CC STATE: Illinois  
CC COUNTRY: USA  
CC ZIP: 60606  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/002,202  
CC FILING DATE: 08-JAN-1993  
CC CLASSIFICATION: 514







FILED DATE: 23 MAR 1992  
ATTORNEY/AGENT INFORMATION:

Query	Query Match	Best Local Similarity	Best Local Similarity	Score	Pred
14:06	14:06	47.18	47.18	14:06	Pred

Query Match	Score	Score 69:	DB 3:	Length	532
-------------	-------	-----------	-------	--------	-----

Query Match : 14.68; Score 69; DB 3; Length 532;  
Best Local Similarity 47.18; Pred. No. 8.54e+01;



Tue Jan 4 11:52:15 2000

US-09-209-961-20.rai

Page 8

Matches 8; Conservative 4; Mismatches 4; Indels 1; Gaps 1;

Db 152 RWLVLLISMVCIITAMI 168

:|:|: | | :|:

Oy 43 KWLLIMTFMAGCR-GMI 58

Search completed: Wed Dec 22 10:27:30 1999  
Job time : 7 secs.

```

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Dec 22 10:26:00 1999: Maspar time 6.68 Seconds
---353.834 Million cell updates/sec
Tabular output not generated.

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Query			DB	ID	Description	Pred. No.
	Score	Match	Length				
1	95	20.2	593	2	S72769	B1496_FL14 protein	1 57e-02
2	94	20.0	591	2	F70871	probable membrane pro	2 24e-02
3	88	18.7	607	2	F69769	conserved hypothetical	1 80e-01
4	84	17.8	362	2	S36766	prostaglandin E recep	6 87e-01
5	84	17.8	395	2	S36765	prostaglandin E recep	6 87e-01
6	84	17.8	397	2	S36767	prostaglandin E recep	6 87e-01
7	84	17.8	417	2	S36764	prostaglandin E recep	6 87e-01
8	82	17.4	198	2	S55499	superoxide dismutase	1 32e+00
9	80	17.0	373	2	A69158	sensory transduction	2 53e+00
10	80	17.0	895	2	S75894	hypothetical protein	2 53e+00
11	79	16.8	60	2	S59092	hypothetical protein	3 48e+00
12	79	16.8	615	1	ABCH5	serum albumin precurs	3 48e+00
13	78	16.6	361	2	A53216	prostaglandin E2 rece	4 77e+00
14	78	16.6	363	2	B64807	abrB protein - Escher	4 77e+00
15	78	16.6	365	2	S51315	prostaglandin E recep	4 77e+00
16	78	16.6	365	2	I38748	prostaglandin receptor	4 77e+00
17	78	16.6	367	2	I02056	prostaglandin E2 rece	4 77e+00
18	78	16.6	374	2	I38747	prostaglandin receptor	4 77e+00
19	78	16.6	374	2	S51317	prostaglandin E recep	4 77e+00
20	78	16.6	388	2	S51316	prostaglandin E recep	4 77e+00
21	78	16.6	388	2	I38750	prostaglandin receptor	4 77e+00
22	78	16.6	390	2	S51313	prostaglandin E recep	4 77e+00
23	78	16.6	390	2	A43375	prostaglandin E recep	4 77e+00

## ALIGNMENTS

RESULTE 1

ENTIRE TITLE ORGANISM DATE
-------------------------------------

ACCESSIONS  
REFERENCE  
#author:  
#submit:  
#descri:  
#access:

##sta  
##mo  
##re:  
##cr

```

##cross-references EMBL:U00013; NID:g466868; PID:g466891
SUMMARY
#length 593 #molecular-weight 63037 #checksum 126

```

```
Query Match      20.2%; Score 95; DB 2; Length 593;
Best Local Similarity 39.0%; Pred. No. 1.57e-02;
Matches 16; Conservative 6; Mismatches 18; Indels 1; Gaps 1;
```

70 LLNLF<sup>3</sup>:I-QTVSLTWTTTCAYVMTLAWLMIGRALRRQM 109  
 ||| : : : : : ||| : : :  
 17 LLTAYIRKPHRTGNHLRTSGGMVTVVKSWLLMTMTWAGCRGM 57

## RESULT

COUNTRY	TITLE

ALTERNATE ORGANISM DATE:	DATE:

ACCESSION  
REFERENCE  
#a:1:bo

01736B4

##journal Nature (1998) 393:537-544  
 ##title Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.

##cross-references GB:AL021184; GB:AL123456; NID:g261498; PID:el237450;

##accession F70871

##status preliminary; nucleic acid sequence not shown;

##molecule\_type DNA

##residues 1-591 ##label COL

##cross-references GB:AL021184; GB:AL123456; NID:g261498; PID:el237450;

##experimental\_source strain H37Rv

GENETICS

##gene

##summary

Query Match 20.08; Score 94; DB 2; Length 591;

Best Local Similarity 39.08; Pred. No. 2.24e-02;

Matches 16; Conservative 6; Mismatches 18; Indels 1; Gaps 1;

Db 70 LNLPSRI-QTVSLTMTTGGVMMALAWMLGRFTLGRMR 109

QY 17 LLLIQLRKPHGTGHLKTSGGVMTWVKWLLMTFMACRCGM 57

RESULT 3

ENTRY conserved hypothetical protein ydaO - Bacillus subtilis

TITLE formal\_name Bacillus subtilis

ORGANISM 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change

DATE 24-Sep-1998

ACCESSIONS F69769

REFERENCE A69580

##authors Kunst, F.; Ogasawara, N.; Mosser, I.; Albertini, A.M.;

Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;

Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,

A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;

Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;

Choi, S.K.; Codani, J.F.; Connerton, I.F.; Cummings, N.J.;

Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;

Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.;

Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,

M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,

S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;

Guisepi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood,

C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;

Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;

Kashara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,

Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.;

Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;

Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;

Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,

M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, R.P.; O'Reilly,

V.; Pohl, T.M.; Portetelle, D.; Porwollik, S.; Prescott,

A.M.; Prescan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;

Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;

Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;

Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;

Sekowska, A.; Seror, S.J.; Serroir, P.; Shin, B.S.; Soldo,

T.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;

Takemaru, K.; Takeuchi, M.; Tanakoshi, A.; Tanaka, T.;

Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;

Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;

Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.;

Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,

K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;

Yoshikawa, H.; Danchin, A.

Nature (1997) 390:249-256

The complete genome sequence of the Gram-positive bacterium

Bacillus subtilis.

##cross-references F69769

##accession

##status preliminary; nucleic acid sequence not shown;

##molecule\_type DNA

##residues 1-607 ##label KUN

##cross-references GB:Z99106; GB:AL009126; NID:g2632653; PID:el182398;

##experimental\_source strain 168

GENETICS

##gene

##summary

Query Match 18.7%; Score 88; DB 2; Length 607;

Best Local Similarity 32.7%; Pred. No. 1.80e-01;

Matches 16; Conservative 10; Mismatches 23; Indels 0; Gaps 0;

Db 181 LLVIAVGLFKLMTGQIDQPAHHTSLGTPVAGITLFLLLKAFSSCSAL 229

QY 9 LGVVKAGLLLIQTRKPHGTGHLKTSGGVMTWVKWLLMTFMACRCGM 57

RESULT 4

ENTRY S36766 #type complete

TITLE prostaglandin E receptor 3C - bovine

ORGANISM #formal\_name Bos primigenius taurus #common\_name cattle

DATE 10-Dec-1993 #sequence\_revision 27-Feb-1997 #text\_change

ACCESSIONS S36766

REFERENCE S36764

##authors Namba, T.; Sugimoto, Y.; Negishi, M.; Irie, A.; Ushikubi, F.;

Kakizuka, A.; Ito, S.; Ichikawa, A.; Narumiya, S.

Nature (1993) 365:166-170

Alternative splicing of C-terminal tail of prostaglandin E

receptor subtype EP3 determines G-protein specificity.

##cross-references MUID:93382504

##accession S36766

##status preliminary

##molecule\_type mRNA

##residues 1-362 ##label NAM

##cross-references GB:D21347; NID:g456444; PID:dl005346; PID:g456445

CLASSIFICATION #superfamily prostaglandin E receptor EPI

SUMMARY #length 362 #molecular-weight 39887 #checksum 8225

Query Match 17.8%; Score 84; DB 2; Length 362;

Best Local Similarity 37.5%; Pred. No. 6.87e-01;

Matches 12; Conservative 8; Mismatches 11; Indels 1; Gaps 1;

Db 156 LATRAPHWSSHMKTSYTRAVLLGVWLAVLAF 187

QY 20 LQTRKPH-TGNHLKTSGGVMTWVKWLLMTF 50

RESULT 5

ENTRY S36765 #type complete

TITLE prostaglandin E receptor 3B - bovine

ORGANISM #formal\_name Bos primigenius taurus #common\_name cattle

DATE 10-Dec-1993 #sequence\_revision 27-Feb-1997 #text\_change

ACCESSIONS S36765

REFERENCE S36764

##authors Namba, T.; Sugimoto, Y.; Negishi, M.; Irie, A.; Ushikubi, F.;

Kakizuka, A.; Ito, S.; Ichikawa, A.; Narumiya, S.

Nature (1993) 365:166-170

Alternative splicing of C-terminal tail of prostaglandin E

receptor subtype EP3 determines G-protein specificity.

##cross-references MUID:93382504

##accession S36765

##status preliminary

##molecule\_type mRNA

##residues 1-385 ##label NAM

##cross-references GB:D21346; NID:g415725; PID:dl005345; PID:g415726

CLASSIFICATION #superfamily prostaglandin E receptor EPI

SUMMARY #length 385 #molecular-weight 42466 #checksum 6243



```

##residues 1-373 #label MTH
##cross-references GB:AE000828; GB:AE000666; NID:g2621504; PID:g2621510
##experimental_source strain Delta H
GENETICS
#gene MTH444
#summary #length 373 #molecular-weight 42263 #checksum 7789
Query Match 17.08; Score 80; DB 2; Length 373;
Best Local Similarity 37.5%; Pred. No. 2.53e+00;
Matches 18; Conservative 6; Mismatches 20; Indels 4; Gaps 2;
Db 76 WGVLVAVFVALLIOPMTAVG--LNISDELRLVL--LLLEFEMVA 119
QY 6 WGHGVVKGANGLLIQRKPHNTSGMVTWVKWLLMTFMAG 53

RESULT 10
ENTRY S75894 #type complete
TITLE hypothetical protein - Synecocystis sp. (strain PCC 6803)
ORGANISM #formal_name Synecocystis sp.
#variety PCC 6803
DATE 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
21-Aug-1998
ACCESSIONS S75894
REFERENCE S74322
#authors Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.;
Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpou,
S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
Yasuda, M.; Tabata, S.
#journal DNA Res. (1996) 3:109-136
#title Sequence analysis of the genome of the unicellular
cyanobacterium Synecocystis sp. PCC6803. II. Sequence
determination of the entire genome and assignment of
potential protein-coding regions.
#cross-references MUD:97061201
#accession S75894
##status preliminary
##molecule_type DNA
##residues 1-885 #label KAN
##cross-references EMBL:D90913; GB:AB001339; NID:gl653348; PID:dl019086;
PID:gl653439
#note the nucleotide sequence was submitted to the EMBL Data
Library, June 1996
SUMMARY #length 885 #molecular-weight 101491 #checksum 6510
Query Match 17.08; Score 80; DB 2; Length 885;
Best Local Similarity 39.5%; Pred. No. 2.53e+00;
Matches 15; Conservative 7; Mismatches 13; Indels 3; Gaps 3;
Db 409 MYGDLNLHOLPIVRAI--LLEMQRKINTGADGIRVDGG 445
QY 1 MYT-SLWGLGVVKGANGLLIQRKPHNTGHNH-LKTSGG 36

RESULT 11
ENTRY S59092 #type complete
TITLE hypothetical protein 60 - red alga (Chondrus crispus)
ORGANISM #formal_name mitochondrion Chondrus crispus #common_name
carragheen
DATE 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change
24-Sep-1998
ACCESSIONS S59092
REFERENCE S59078
#authors Leblanc, C.; Boyen, C.; Richard, O.; Bonnard, G.;
Grienerberger, J.M.; Kloareg, B.
#journal J. Mol. Biol. (1995) 250:484-495
#title Complete sequence of the mitochondrial DNA of the rhodophyte
Chondrus crispus (Gigartinales). Gene content and genome
organization.
#cross-references MUD:95341681

```

```

#accession S59092
##status S59092 #nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues 1-60 #label LEB
##cross-references EMBL:247547; NID:gl019057; PID:gl019071
##experimental_source female gametophytes
#note the nucleotide sequence was submitted to the EMBL Data
Library, January 1995
GENETICS
#genome mitochondrion
#genetic_code SGC3
KEYWORDS mitochondrion
SUMMARY #length 60 #molecular-weight 6633 #checksum 6517
Query Match 16.8%; Score 79; DB 2; Length 60;
Best Local Similarity 42.9%; Pred. No. 3.48e+00;
Matches 6; Conservative 7; Mismatches 1; Indels 0; Gaps 0;
Db 1 MVRHWLMLAFKGS 14
QY 40 MVRKWLMTFMAG 53

RESULT 12
ENTRY ABCHS #type complete
TITLE serum albumin precursor - chicken
ORGANISM #formal_name Gallus gallus #common_name chicken
DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
05-Sep-1997
ACCESSIONS S15571; A05078; A13451
REFERENCE S15571
#authors Cassidy, A.I.; Salkild, C.K.; Baverstock, P.; Wallace, J.C.
#submission submitted to the EMBL Data Library, July 1991
#accession S15571
#molecule_type mRNA
##residues 1-615 #label CAS
##cross-references EMBL:X60688; NID:g63747; PID:g63748
#accession A05078
#authors Hache, R.J.G.; Wiskocil, R.; Vasa, M.; Roy, R.N.; Lau,
P.C.K.; Deeley, R.G.
#journal J. Biol. Chem. (1983) 258:4556-4564
#title The 5' noncoding and flanking regions of the avian very low
density apolipoprotein II and serum albumin genes.
Homologies with the egg white protein genes.
#cross-references MUD:83161037
#accession A05078
#molecule_type DNA
##residues 1-28 #label HAC
##cross-references GB:V00381; NID:g63038; PID:g63039
#accession A13451
#authors Rosen, A.M.; Geller, D.M.
#journal Biochem. Biophys. Res. Commun. (1977) 78:1060-1066
#title Chicken microsomal albumin: amino terminal sequence of
chicken proalbumin.
#cross-references MUD:78019943
#accession A13451
#molecule_type protein
##residues 19-23, 'M', 25-30 #label ROS
COMMENT Serum albumin is synthesized in the liver as preproalbumin. It
binds copper, nickel, calcium (weakly, at 2-3 sites), and
protoporphyrin, long-chain fatty acids, prostaglandins, steroid
hormones (weak bonds with these hormones promote their transfer
across the membranes), thyroxine, and triiodothyronine.
CLASSIFICATION #superfamily serum albumin; serum albumin repeat homology
carrier protein; duplication; metal binding; plasma
KEYWORDS
FEATURE
1-18 #domain signal sequence #status predicted #label SIG\
19-26 #domain propeptide #status predicted #label PRO\
27-613 #product serum albumin #status predicted #label MAT\
32-206 #domain serum albumin repeat homology #label SAL\
225-398 #domain serum albumin repeat homology #label SA2\
417-596 #domain serum albumin repeat homology #label SA3\
30 #binding_site copper (His) #status predicted\

```

```
80-89,102-118,
117-128,152-197,
196-205,228-274,
273-281,293-307,
306-317,344-389,
388-397,420-466,
465-476,489-505,
504-515,542-587,
586-595
SUMMARY #length 615 #molecular-weight 69918 #checksum 4822
Query Match 16.8%; Score 79; DB 1; Length 615;
Best Local Similarity 46.7%; Pred. No. 3.48e+00;
Matches 14; Conservative 6; Mismatches 7; Indels 3; Gaps 3;

Db 556 LLINIKRPQTEQIKTIADGFTAMVDK 585
QY 17 LLI-LQTRKPH-TGNHKT-SGGWMTMVK 43

RESULT 13
ENTRY A53216 #type complete
TITLE prostaglandin E2 receptor, subtype EP3 (clone 72A) - rabbit
ORGANISM #formal_name Oryctolagus cuniculus #common_name domestic
DATE 19-Oct-1995 #sequence_revision 19-Oct-1995 #text_change
ACCESSIONS A53216
REFERENCE A53216
#authors Breyer, R.M.; Emeson, R.B.; Tarng, J.L.; Breyer, M.D.; Davis,
L.S.; Abramson, R.M.; Ferrenbach, S.M.
#journal J. Biol. Chem. (1994) 269:6163-6169
#title Alternative splicing generates multiple isoforms of a rabbit
prostaglandin E-2 receptor.
#cross-references MUID:94164982
#accession A53216
#status preliminary
#molecule_type mRNA
#residues 1-361 #label BRE
#cross-references GB:U04273; NID:G467978; PID:G467979
CLASSIFICATION #superfamily prostaglandin E receptor Epl
KEYWORDS #alternative splicing; transmembrane protein
SUMMARY #length 361 #molecular-weight 40011 #checksum 4634
Query Match 16.6%; Score 78; DB 2; Length 361;
Best Local Similarity 31.3%; Pred. No. 4.77e+00;
Matches 10; Conservative 9; Mismatches 12; Indels 1; Gaps 1;

Db 153 LAIRAPHYASHMKTRATRAVLLGWLAVLAF 184
QY 20 LQTRKPH-TGNHKTSGGWMTMVKWLLMTF 50

RESULT 14
ENTRY B64807 #type complete
TITLE abrB protein - Escherichia coli
ORGANISM #formal_name Escherichia coli
DATE 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
ACCESSIONS B64807
REFERENCE A64720
#authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
Y.
#journal Science (1997) 277:1453-1462
#title The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession B64807
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-363 #label BLAT
```

```
##cross-references GB:AE000174; GB:U00096; NID:q1786920; PID:q1786933;
##experimental_source strain K-12, substrain MGL155
GENETICS
#gene abrB
#keywords transmembrane protein
FEATURE
22-38 #domain transmembrane #status predicted #label TM1\
40-56 #domain transmembrane #status predicted #label TM2\
98-114 #domain transmembrane #status predicted #label TM3\
162-178 #domain transmembrane #status predicted #label TM4\
198-214 #domain transmembrane #status predicted #label TM5\
277-293 #domain transmembrane #status predicted #label TM6\
329-345 #domain transmembrane #status predicted #label TM7\
SUMMARY #length 363 #molecular-weight 38939 #checksum 8760
Query Match 16.6%; Score 78; DB 2; Length 363;
Best Local Similarity 40.7%; Pred. No. 4.77e+00;
Matches 11; Conservative 7; Mismatches 7; Indels 2; Gaps 2;

Db 228 AGAVLQ-SGLITIELPEWLLAMAYMA 253
QY 27 TGNHLKTSGGWMTM-VKKWLLMTFMA 52

RESULT 15
ENTRY S51315 #type complete
TITLE prostaglandin E receptor, subtype EP3 splice form b - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 07-May-1995 #sequence_revision 01-Sep-1995 #text_change
ACCESSIONS S68995; S41078; S51315
REFERENCE S68994
#authors Schmid, A.; Thierach, K.H.; Schleuning, W.D.; Dinter, H.
#journal Eur. J. Biochem. (1995) 228:33-30
#title Splice variants of the human EP(3) receptor for prostaglandin
E(2).
#cross-references MUID:95188908
#accession S68995
#molecule_type mRNA
#residues 1-365 #label SC2
#cross-references EMBL:X83859; NID:G633211; PID:G633212
#note Submitted to the EMBL Data Library, January 1995
REFERENCE S41076
#authors Adam, M.; Boie, Y.; Rushmore, T.H.; Mueller, G.; Bastien, L.;
McKee, K.T.; Metters, K.M.; Abramovitz, M.
#journal FEBS Lett. (1994) 338:170-174
#title Cloning and expression of three isoforms of the human EP(3)
prostanoid receptor.
#cross-references MUID:94139918
#accession S41073
#status preliminary
#molecule_type mRNA
#residues 1-365 #label ADA
GENETICS
#gene GDB:PTGER3
#cross-references GDB:321928; OMIM:176806
#map_position lp31.2-lp31.2
CLASSIFICATION #superfamily prostaglandin E receptor Epl
KEYWORDS #alternative splicing
SUMMARY #length 365 #molecular-weight 40501 #checksum 2560
Query Match 16.8%; Score 78; DB 2; Length 365;
Best Local Similarity 31.3%; Pred. No. 4.77e+00;
Matches 10; Conservative 9; Mismatches 12; Indels 1; Gaps 1;

Db 157 LAIRAPHYASHMKTRATRAVLLGWLAVLAF 188
QY 20 LQTRKPH-TGNHLKTSGGWMTMVKWLLMTF 50
Search completed: Wed Dec 22 10:26:09 1999
```

Tue Jan 4 11:52:17 2000

US-09-209 961-20 rpr

Page 6

Job time : 9 secs.

\*\*\*\*\*  
WIREHIA  
\*\*\*\*\*  
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Wed Dec 22 10:26:27 1999; Maspar time 4.57 Seconds  
Tabular output not generated. 364.668 Million cell updates/sec

Title: >US-09-209-961-20  
Description: (1-59) from US09209961.pep  
Perfect Score: 471  
Sequence: 1 MVTSLWGLGVKANGILLIL.....MVKWLLILTFMAGCRGMIV 59

Scoring table: PAM 150  
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot37  
1:swissprot

Statistics: Mean 37.670; Variance 59.374; scale 0.634

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	84	17.8	417	1 PER3_BOVIN	PROSTAGLANDIN E2 RECP	1.76e-01
2	83	17.6	373	1 PER3_PIG	PROSTAGLANDIN E2 RECP	2.58e-01
3	81	17.2	508	1 NO60_DROME	NUCLEOLAR PROTEIN AT B	5.28e-01
4	79	16.8	615	1 ABRU_CHICK	SERUM ALBUMIN PRECURSOR	1.08e+00
5	78	16.6	363	1 ABRU_ECOLI	ABRB PROTEIN	1.53e+00
6	78	16.6	390	1 PER3_HUMAN	PROSTAGLANDIN E2 RECP	1.53e+00
7	78	16.6	411	1 PER3_RABIT	PROSTAGLANDIN E2 RECP	1.53e+00
8	78	16.6	547	1 MERA_STAAR	MERCURIC REDUCTASE (EC	1.53e+00
9	78	16.6	672	1 BGAL_BACST	BETA-GALACTOSIDASE I (	1.53e+00
10	76	16.1	415	1 AMSJ_ERWAM	AMYLOVORAN BIOSYNTHESI	3.07e+00
11	76	16.1	564	1 AGGL_RICCO	AGGLUTININ PRECURSOR (	3.07e+00
12	76	16.1	811	1 PIRA_RHORI	PRIMOSOMAL PROTEIN N'	3.07e+00
13	75	15.9	93	1 YSL1_THEFL	HYPOTHETICAL PROTEIN I	4.32e+00
14	75	15.9	152	1 UL04_HCMVA	EARLY GLYCOPROTEIN GP4	4.32e+00
15	75	15.9	398	1 Y05P_MYCTU	HYPOTHETICAL 43.7 KD P	4.32e+00
16	75	15.9	783	1 FUR1_XENLA	FURIN 1 PRECURSOR (EC	4.32e+00
17	75	15.9	1179	1 VGS6_HSV11	HYPOTHETICAL GENE 56 P	4.32e+00
18	74	15.7	209	1 YC02_KLEPN	3-HYDROXYANTHRANILATE	6.06e+00
19	74	15.7	286	1 XHAO_HUMAN	3-HYDROXYANTHRANILATE	6.06e+00
20	74	15.7	317	1 APE_PAPAN	APOLIPOPROTEIN E PRECU	6.06e+00
21	74	15.7	317	1 APE_MACFA	APOLIPOPROTEIN E PRECU	6.06e+00
22	74	15.7	317	1 APE_HUMAN	APOLIPOPROTEIN E PRECU	6.06e+00
23	74	15.7	519	1 NIFL_AZOVI	NITROGEN FIXATION REGU	6.06e+00

24	74	15.7	576	1 RICI_RICCO	RICIN PRECURSOR (RRNA	6.06e+00
25	74	15.7	773	1 CPTI_HUMAN	MITOCHONDRIAL CARNITIN	6.06e+00
26	74	15.7	971	1 SN2L_CABEL	POSSIBLE GLOBAL TRANS	6.06e+00
27	73	15.5	102	1 DAPI_HUMAN	DEATH-ASSOCIATED PROTE	8.47e+00
28	73	15.5	104	1 VPR_HV2ST	VPR PROTEIN (R ORF PRO	8.47e+00
29	73	15.5	180	1 RBS_MUSAC	RIBULOSE BISPHOSPHATE	8.47e+00
30	73	15.5	434	1 ENO_LOLPE	ENOLASE (EC 4.2.1.11)	8.47e+00
31	73	15.5	693	1 YETI_SCHPO	PUTATIVE ABC TRANSPORT	8.47e+00
32	73	15.5	1133	1 SREI_CRIGR	STEROL REGULATORY ELEM	8.47e+00
33	72	15.3	135	1 YIDQ_ECOLI	HYPOTHETICAL 14.8 KD P	1.18e+01
34	72	15.3	269	1 CDSA_BACSU	PHOSPHATIDATE CYTIDILY	1.18e+01
35	72	15.3	399	1 FDBB_METFO	FORMATE DEHYDROGENASE	1.18e+01
36	72	15.3	584	1 PME_BRANA	PROBABLE PECTINESTERAS	1.18e+01
37	72	15.3	590	1 SYD_ECOLI	ASPARTYL-TRNA SYNTHETA	1.18e+01
38	72	15.3	618	1 Y4RQ_RHISN	HYPOTHETICAL 69.3 KD P	1.18e+01
39	72	15.3	645	1 COX1_PARTE	CYTOCHROME C OXIDASE P	1.18e+01
40	71	15.1	237	1 FGF3_XENLA	FIBROBLAST GROWTH FACT	1.63e+01
41	71	15.1	529	1 ACH2_HUMAN	NEURONAL ACETYLCHOLINE	1.63e+01
42	71	15.1	1027	1 YDIJ_HAEIN	HYPOTHETICAL PROTEIN H	1.63e+01
43	71	15.1	1227	1 B3A3_RAT	ANION EXCHANGE PROTEIN	1.63e+01
44	71	15.1	1515	1 GDE_HUMAN	GLYCOGEN DEBRANCHING E	1.63e+01
45	71	15.1	2594	1 7LES_DROVI	SEVENLESS PROTEIN (EC	1.63e+01

ALIGNMENTS

RESULT 1  
ID PER3\_BOVIN STANDARD; PRT; 417 AA.  
AC P34979;  
DT 01-FEB-1994 (REL. 28, CREATED)  
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE PROSTAGLANDIN E2 RECEPTOR, EP3 SUBTYPE (PROSTANOID EP3 RECEPTOR) (PGE  
DE RECEPTOR, EP3 SUBTYPE)  
GN PTGER3.  
OS BOS TAURUS (BOVIN)  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC ARTIODACTYLA; RUMINANTIA; PECORA; BOVIDAE; BOVINA; BOS.  
[1]  
RP SEQUENCE FROM N.A.  
KC TISSUE-ADRENAL MEDULLA;  
IX MEDLINE; 53382504.  
RA NAKAZUKA A., ITO S., ICHIKAWA A., NARUMIYA S.S.;  
ET "Alternat-ve splicing of C-terminal tail of prostaglandin E receptor  
RT subtype EP3 determines G-protein specificity";  
RL NATURE 365:166-170(1993).  
CC -!- FUNCTION: RECEPTOR FOR PROSTAGLANDIN E2 (PGE2); THE EP3 RECEPTOR  
MAY BE INVOLVED IN INHIBITION OF GASTRIC ACID SECRETION,  
MODULATION OF NEUROTRANSMITTER RELEASE IN CENTRAL AND PERIPHERAL  
NEURONS. INHIBITION OF SODIUM AND WATER REABSORPTION IN KIDNEY  
TUBULUS AND CONTRACTION IN UTERINE SMOOTH MUSCLE. THE ACTIVITY OF  
THIS RECEPTOR CAN COUPLE TO BOTH THE INHIBITION OF ADENYLATE  
CYCLASE MEDIATED BY G-I PROTEINS, AND TO AN ELEVATION OF  
INTRACELLULAR CALCIUM. THE VARIOUS FORMS INTERACT WITH DIFFERENT  
SECOND MESSENGER SYSTEMS.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -!- ALTERNATIVE PRODUCTS: AT LEAST FOUR FORMS (A TO D) THAT DIFFERS  
AT THE C-TERMINUS ARE PRODUCED BY ALTERNATIVE SPLICING OF THIS  
RECEPTOR GENE. THEY HAVE IDENTICAL LIGAND BINDING PROPERTIES BUT  
DIFFERENT COUPLING PROPERTIES WITH G PROTEINS: EP3A COUPLES TO  
G(1)/G(0) PROTEINS; EP3B AND EP3C COUPLE TO G(S), AND EP3D COUPLES  
TO G(1), G(S) AND G(P).  
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----



DR EMBL; D21345; G415724; --  
DR EMBL; D21346; G415726; --  
DR EMBL; D21347; G456445; --  
DR EMBL; D21348; G415730; --  
DR PIR; S36764; S36764.  
DR GCRDB; GCR\_0784; --  
DR GCRDB; GCR\_0785; --  
DR GCRDB; GCR\_0786; --  
DR GCRDB; GCR\_0787; --  
DR PROSITE; PS00237; G\_PROTEIN\_RECEPTOR; FALSE\_NEG.  
DR PFAM; PF00001; 7tm\_1; 1.  
KW G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN;  
KW LIPOPROTEIN; PALMITATE; PHOSPHORYLATION; ALTERNATIVE SPLICING.  
FT DOMAIN 1 52  
FT TRANSSEM 53 77  
FT DOMAIN 78 90  
FT TRANSSEM 91 111  
FT DOMAIN 112 130  
FT TRANSSEM 131 152  
FT DOMAIN 153 173  
FT TRANSSEM 175 196  
FT DOMAIN 197 226  
FT TRANSSEM 227 252  
FT DOMAIN 253 282  
FT TRANSSEM 283 306  
FT DOMAIN 307 326  
FT TRANSSEM 327 348  
FT DOMAIN 349 417  
FT VARSPLIC 359 387  
FT VARSPLIC 388 417  
FT VARSPLIC 359 362  
FT VARSPLIC 363 417  
FT VARSPLIC 359 385  
FT VARSPLIC 386 417  
FT CARBOHYD 17 17  
FT CARBOHYD 35 35  
FT CARBOHYD 216 216  
FT CARBOHYD 307 307  
SQ SEQUENCE 417 AA; 46362 MW; 5AB01C72 CRC32;  
  
Query Match 17.8%; Score 84; DB 1; Length 417;  
Best Local Similarity 37.5%; Pred. No. 1.76e-01;  
Matches 12; Conservative 8; Mismatches 11; Indels 1; Gaps 1;  
  
Db 156 LATRAPHWYSSHMTSVTRAVLLGVWLVAVLAF 187  
QY 20 LQTRKPH-TGNHLKTSVGMVTKWLLMTF 50  
| | | | | : : : : :  
PRT; 373 AA.  
STANDARD; PRT; 508 AA.  
RESULTS 3  
ID NO60\_DROME STANDARD; PRT; 508 AA.  
AC 044081;  
DT 15-DEC-1998 (REL. 37, CREATED)  
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)  
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
DE NUCLEOLAR PROTEIN AT BAND 60B.  
GN NOP60B.  
OS DROSOPHILA MELANOGASTER (FRUIT FLY).  
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;  
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;  
OC DROSOPHILIDAE; DROSOPHILA.  
RN [1]  
RP SEQUENCE FROM N.A.  
  
Query Match 17.6%; Score 83; DB 1; Length 373;  
Best Local Similarity 34.4%; Pred. No. 2.55e-01;  
Matches 11; Conservative 9; Mismatches 11; Indels 1; Gaps 1;  
  
Db 134 LATRAPHWYSSHMTSATRAVLLGVWLVAVLAF 165  
QY 20 LQTRKPH-TGNHLKTSVGMVTKWLLMTF 50  
| | | | | : : : : :  
PRT; 373 AA.  
STANDARD; PRT; 508 AA.  
RESULTS 3  
ID NO60\_DROME STANDARD; PRT; 508 AA.  
AC 044081;  
DT 15-DEC-1998 (REL. 37, CREATED)  
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)  
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
DE NUCLEOLAR PROTEIN AT BAND 60B.  
GN NOP60B.  
OS DROSOPHILA MELANOGASTER (FRUIT FLY).  
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;  
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;  
OC DROSOPHILIDAE; DROSOPHILA.  
RN [1]  
RP SEQUENCE FROM N.A.  
  
CC -1- FUNCTION: RECEPTOR FOR PROSTAGLANDIN E2 (PGE2); THE EP3 RECEPTOR  
CC MAY BE INVOLVED IN INHIBITION OF GASTRIC ACID SECRETION,  
CC MODULATION OF NEUROTRANSMITTER RELEASE IN CENTRAL AND PERIPHERAL  
CC NEURONS, INHIBITION OF SODIUM AND WATER REABSORPTION IN KIDNEY  
CC TUBULUS AND CONTRACTION IN UTERINE SMOOTH MUSCLE. THE ACTIVITY OF  
CC THIS RECEPTOR CAN COUPLE TO BOTH THE INHIBITION OF ADENYLATE  
CC CYCLASE MEDIATED BY G-I PROTEINS, AND TO AN ELEVATION OF  
CC INTRACELLULAR CALCIUM. THE VARIOUS FORMS CAN INTERACT WITH  
CC DIFFERENT SECOND MESSENGER SYSTEMS (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U27083; G984358;  
CC EMBL; U30374; G984362;  
CC GCRDB; GCR\_1576;  
CC PROSITE; PS00237; G\_PROTEIN\_RECEPTOR; FALSE\_NEG.  
CC PFAM; PF00001; 7tm\_1; 1.  
KW G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN;  
KW ALTERNATIVE SPLICING.  
FT DOMAIN 1 30  
FT TRANSSEM 31 55  
FT DOMAIN 56 68  
FT TRANSSEM 69 89  
FT DOMAIN 90 108  
FT TRANSSEM 109 130  
FT DOMAIN 131 151  
FT TRANSSEM 153 174  
FT DOMAIN 175 204  
FT TRANSSEM 205 230  
FT DOMAIN 231 260  
FT TRANSSEM 261 284  
FT DOMAIN 285 304  
FT TRANSSEM 305 326  
FT DOMAIN 327 373  
FT CARBOHYD 13 13  
FT CARBOHYD 194 194  
FT DISULFID 107 185  
FT VARSPLIC 337 373  
SQ SEQUENCE 373 AA; 41416 MW; BFECCOD9 CRC32;  
  
Query Match 17.6%; Score 83; DB 1; Length 373;  
Best Local Similarity 34.4%; Pred. No. 2.55e-01;  
Matches 11; Conservative 9; Mismatches 11; Indels 1; Gaps 1;  
  
Db 134 LATRAPHWYSSHMTSATRAVLLGVWLVAVLAF 165  
QY 20 LQTRKPH-TGNHLKTSVGMVTKWLLMTF 50  
| | | | | : : : : :  
PRT; 373 AA.  
STANDARD; PRT; 508 AA.  
RESULTS 3  
ID NO60\_DROME STANDARD; PRT; 508 AA.  
AC 044081;  
DT 15-DEC-1998 (REL. 37, CREATED)  
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)  
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
DE NUCLEOLAR PROTEIN AT BAND 60B.  
GN NOP60B.  
OS DROSOPHILA MELANOGASTER (FRUIT FLY).  
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;  
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;  
OC DROSOPHILIDAE; DROSOPHILA.  
RN [1]  
RP SEQUENCE FROM N.A.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

	1	18	
KW			PLASMA: METAL-BINDING; LIPID-BINDING; ALBUMIN; REPEAT; SIGNAL;
FT			COPPER.
FT	SIGNAL	1	
FT	PROPEP	19	
FT	CHAIN	24	SERUM ALBUMIN.
FT	REPEAT	31	1.
FT	REPEAT	225	2.
FT	REPEAT	417	3.
FT	METAL	30	COPPER (BY SIMILARITY)

FT	DISULEP	80	89	BY SIMILARITY.
FT	DISULEP	102	118	BY SIMILARITY.
FT	DISULEP	117	128	BY SIMILARITY.
FT	DISULEP	152	197	BY SIMILARITY.
FT	DISULEP	196	205	BY SIMILARITY.
FT	DISULEP	228	274	BY SIMILARITY.
FT	DISULEP	273	281	BY SIMILARITY.
FT	DISULEP	293	307	BY SIMILARITY.
FT	DISULEP	293	321	BY SIMILARITY.

DI	DISULFID	344	369	BY SIMILARITY.
FT	DISULFID	388	397	BY SIMILARITY.
FT	DISULFID	420	466	BY SIMILARITY.
FT	DISULFID	465	476	BY SIMILARITY.
FT	DISULFID	489	505	BY SIMILARITY.
FT	DISULFID	504	515	BY SIMILARITY.
FT	DISULFID	542	587	BY SIMILARITY.
FT	DISULFID	586	595	BY SIMILARITY.
FT	CARBONYD	500	500	POTENTIAL.
FT	CONFLICT	24	24	F -> M (IN REF. 3).
SQ	SEQUENCE	615 AA;	69918 MW;	DC586EEA CRC32;

```

Query Match: 16.8%; Score 79; DB 1; Length 615;
Best Local Similarity 46.7%; Pred. No. 1.08e+00;
Matches 14; Conservative 6; Mismatches 7; Indels 3; Gaps 3;

Db 556 LLINLIRKKQMTTEEQIKTIADGFTAMVDK 585
    ||| ||| : ||| : ||| : |||
Oy 17 LLI-LQTRKPH-TGNHLKT-SGGMTVMYKK 43

RESULT 5
ID ABRB_ECOLI STANDARD; PRG 363 AA.
AC P75747;
DT 01-NOV-1997 (REF. 35. CREATED)

```

DT	01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT	15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DT	ABRB PROTEIN.
CN	ABRB.
OS	ESCHERICHIA COLI.
OC	BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
CC	ESCHERICHIA.
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-K12 / MG1655;
RA	MEDLINE; 97426617.
RA	BLATTNER E.R.; PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA	RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA	BREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA	MAU B., SHAO Y.;
RT	"the complete genome sequence of <i>Escherichia coli</i> K-12.";
KL	SCIENCE 277:1453-1474(1997).

\*The complete genome sequence of *Escherichia coli* K-12.\*;  
SCIENCE 277:1453-1474(1997).

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AE000174; GI786933;  
 CC ECOGENE: EG13310; ABRB.  
 CC TRANSMEMBRANE

FT TRANSMEM 35 55 POTENTIAL.  
 FT TRANSMEM 74 94 POTENTIAL.  
 FT TRANSMEM 98 118 POTENTIAL.  
 FT TRANSMEM 153 173 POTENTIAL.  
 FT TRANSMEM 188 208 POTENTIAL.  
 FT TRANSMEM 211 231 POTENTIAL.  
 FT TRANSMEM 236 256 POTENTIAL.  
 FT TRANSMEM 274 294 POTENTIAL.  
 FT TRANSMEM 303 323 POTENTIAL.  
 FT TRANSMEM 330 350 POTENTIAL.  
 SQ SEQUENCE 363 AA; 38939 MW; BA34DC91 CRC32;

Query Match 16.6%; Score 78; DB 1; Length 363;

Best Local Similarity 40.7%; Pred. NO. 1.53e+00;

Matches 11; Conservative 7; Mismatches 7; Indels 2; Gaps 2;

DB 228 AGAVLO-SGQITIELPEWLLAMAYMA 253

QY 27 TGNHLKTSGGVMTM-VKKWLLMTFMA 52

# RESULT

ID PER3 HUMAN STANDARD; PRT; 390 AA.  
 AC P43115; Q12943; Q12944; Q12945; Q16546;  
 DT 01-NOV-1995 (REL. 32, CREATED)  
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
 DE PROSTAGLANDIN E2 RECEPTOR, EP3 SUBTYPE (PROSTANOID EP3 RECEPTOR) (PGE  
 DE RECEPTOR, EP3 SUBTYPE).  
 GN PIGR3.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
 [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 94139918.  
 RA ADAM M., BOYE Y., RUSHMORE T.H., MULLER G., BASTIEN L., MCKEE K.T.,  
 RA METTERS K.M., ABRAMOVITZ M.;  
 RT "Cloning and expression of three isoforms of the human EP3 prostanoid  
 RT receptor".  
 RL FEBS LETT. 338:170-174(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-UTERUS;  
 RX MEDLINE; 95188908.  
 RA SCHMID A., THIERAUCH K.H., SCHLEUNING W.D., DINTER H.;  
 RT "Splice variants of the human EP3 receptor for prostaglandin E2".  
 RL EUR. J. BIOCHEM. 228:23-30(1995).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-KIDNEY;  
 RX MEDLINE; 94161771.  
 RA YANG J., XIA M., GOETZL E.J., AN S.;  
 RT "Cloning and expression of the EP3-subtype of human receptors for  
 RT prostaglandin E2".  
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 198:999-1006(1994).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 94183149.  
 RA KUNAPULI S.P., FEN MAO G., BASTEPE M., LIU-CHEN L.-Y., LI S.,

RA CHEUNG P.P., DERIEL J.K., ASHEY B.;  
 RT "Cloning and expression of a prostaglandin E receptor EP3 subtype  
 RT from human erythroleukemia cells".  
 RL BIOCHEM. J. 298:263-267(1994).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-UTERUS;  
 RX MEDLINE; 95072021.  
 RA AN S., YANG J., SO S.W., ZENG L., GOETZL E.J.;  
 RT "Isoforms of the EP3 subtype of human prostaglandin E2 receptor  
 RT transduce both intracellular calcium and cAMP signals".  
 RL BIOCHEMISTRY. 33:14496-14502(1994).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-KIDNEY;  
 RX MEDLINE; 96074809.  
 RA KOTANI M., TANAKA I., OGAWA Y., USUI T., MORI K., ICHIKAWA A.,  
 RA NARUMIYA S., YOSHIMI T., NAKAO K.;  
 RT "Molecular cloning and expression of multiple isoforms of human  
 RT prostaglandin E receptor EP3 subtype generated by alternative  
 RT messenger RNA splicing: multiple second messenger systems and  
 RT tissue-specific distributions".  
 RL MOL. PHARMACOL. 48:869-879(1995).  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-SMALL INTESTINE;  
 RX MEDLINE; 94356288.  
 RA REGAN J.W., BAILEY T.J., DONELLO J.E., PIERCE K.L., PEPPERL D.J.,  
 RA ZHANG D., KEDZIE K.M., FAIRBAIRN C.E., BOGARDUS A.M., WOODWARD D.F.,  
 RA GIL D.W.;  
 RT "Molecular cloning and expression of human EP3 receptors: evidence of  
 RT three variants with differing carboxyl termini".  
 RL BR. J. PHARMACOL. 112:377-385(1994).  
 RN [8]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 97230456.  
 RA KOTANI M., TANAKA I., OGAWA Y., USUI T., TAMURA N., MORI K.,  
 RA NARUMIYA S., YOSHIMI T., NAKAO K.;  
 RT "Structural organization of the human prostaglandin EP3 receptor  
 RT subtype gene (PIGR3)".  
 RL GENOMICS 40:425-434(1997).  
 CC -1- FUNCTION: RECEPTOR FOR PROSTAGLANDIN E2 (PGE2); THE EP3 RECEPTOR  
 CC MAY BE INVOLVED IN INHIBITION OF GASTRIC ACID SECRETION,  
 CC MODULATION OF NEUROTRANSMITTER RELEASE IN CENTRAL AND PERIPHERAL  
 CC NEURONS, INHIBITION OF SODIUM AND WATER REABSORPTION IN KIDNEY  
 CC TUBULUS AND CONTRACTION IN UTERINE SMOOTH MUSCLE. THE ACTIVITY OF  
 CC THIS RECEPTOR CAN COUPLE TO BOTH THE INHIBITION OF ADENYLATE  
 CC CYCLASE MEDIATED BY G-I PROTEINS, AND TO AN ELEVATION OF  
 CC INTRACELLULAR CALCIUM. THE VARIOUS FORMS CAN INTERACT WITH  
 CC DIFFERENT SECOND MESSENGER SYSTEMS (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED AT LEAST IN SMALL INTESTINE, HEART,  
 CC AND PANCREAS.  
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST SIX FORMS THAT DIFFERS AT THE C-  
 CC TERMINUS ARE PRODUCED BY ALTERNATIVE SPLICING OF THIS RECEPTOR  
 CC GENE. THEY HAVE IDENTICAL LIGAND BINDING PROPERTIES BUT DIFFERENT  
 CC COUPLING PROPERTIES WITH G PROTEINS. THE FORMS SHOWN HERE IS KNOWN  
 CC AS EP3-I, EP3A, OR EP3AL.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
 CC -1- CAUTION: IN REF.7 EP3B IS KNOWN AS EP3E, EP3C AS EP3D AND EP3D AS  
 CC EP3F.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: S69200; G545304;  
 DR EMBL: L27490; G440314;  
 DR EMBL: L27488; G440310;

DR EMBL; L27489; G440312; -  
 DR EMBL; X83857; G633208; -  
 DR EMBL; X83858; G633210; -  
 DR EMBL; X83859; G633212; -  
 DR EMBL; X83860; G633214; -  
 DR EMBL; X83861; G633216; -  
 DR EMBL; X83862; G633218; -  
 DR EMBL; X83863; G633220; -  
 DR EMBL; L26976; G483352; -  
 DR EMBL; S69326; G545919; -  
 DR EMBL; L32660; G857668; -  
 DR EMBL; L32661; G484162; -  
 DR EMBL; L32662; G484164; -  
 DR EMBL; D38297; G1199759; -  
 DR EMBL; D38298; G1199765; -  
 DR EMBL; D38299; G1199767; -  
 DR EMBL; D38300; G1199761; -  
 DR EMBL; D38301; G1199763; -  
 DR EMBL; U13214; G532738; -  
 DR EMBL; U13215; G532740; -  
 DR EMBL; U13216; G532742; -  
 DR EMBL; U13217; G532744; -  
 DR EMBL; U13218; G532746; -  
 DR EMBL; D86096; D1020741; JOINED.  
 DR EMBL; D86087; D1020741; JOINED.  
 DR EMBL; D86088; D1020741; JOINED.  
 DR GCRDB; GCR\_0837; -  
 DR GCRDB; GCR\_0886; -  
 DR GCRDB; GCR\_0887; -  
 DR GCRDB; GCR\_0888; -  
 DR GCRDB; GCR\_1279; -  
 DR GCRDB; GCR\_1288; -  
 DR GCRDB; GCR\_1809; -  
 DR GCRDB; GCR\_1810; -  
 DR GCRDB; GCR\_1811; -  
 DR GCRDB; GCR\_1812; -  
 DR GCRDB; GCR\_1813; -  
 DR GCRDB; GCR\_1866; -  
 DR GCRDB; GCR\_1867; -  
 DR GCRDB; GCR\_1868; -  
 DR GCRDB; GCR\_2037; -  
 DR GCRDB; GCR\_2038; -  
 DR GCRDB; GCR\_2039; -  
 DR GCRDB; GCR\_2041; -  
 DR GCRDB; GCR\_2099; -  
 DR GCRDB; GCR\_2100; -  
 DR GCRDB; GCR\_2101; -  
 DR GCRDB; GCR\_2102; -  
 DR GCRDB; GCR\_2103; -  
 DR GCRDB; GCR\_2104; -  
 DR GCRDB; GCR\_2105; -  
 DR GCRDB; GCR\_2650; -  
 DR GCRDB; GCR\_2651; -  
 DR GCRDB; GCR\_2652; -  
 DR GCRDB; GCR\_2653; -  
 DR GCRDB; GCR\_2654; -  
 DR GCRDB; GCR\_2655; -  
 DR GCRDB; GCR\_2656; -  
 DR GCRDB; GCR\_2657; -  
 DR GCRDB; GCR\_2658; -  
 DR MIM; 176806; -  
 DR PROSITE; PS00237; G\_PROTEIN\_RECEPTOR; FALSE\_NEG.  
 DR PFAM; PF00001; 7tm\_1; 1.  
 KW G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN;  
 KW LIPOPROTEIN; PALMITATE; PHOSPHORYLATION; ALTERNATIVE SPLICING.  
 FT DOMAIN 1 53  
 FT TRANSMEM 54 78  
 FT DOMAIN 79 91  
 FT TRANSMEM 92 112  
 FT DOMAIN 113 131  
 FT TRANSMEM 132 153  
 FT DOMAIN 154 174  
 FT TRANSMEM 154 174  
 FT TRANSMEM 174 197  
 FT TRANSMEM 176 197

HP DOMAIN 198 227  
 FT TRANSMEM 228 253  
 FT DOMAIN 254 283  
 FT TRANSMEM 284 307  
 FT DOMAIN 308 327  
 FT TRANSMEM 328 349  
 FT DOMAIN 350 390  
 FT CARBOHYD 18 18  
 FT CARBOHYD 36 36  
 FT VARSPLIC 360 390  
 FT VARSPLIC 360 390  
 FT VARSPLIC 360 390  
 FT VARSPLIC 360 390  
 Note. remainder of annotations omitted.  
 Query Match 16.6%; Score 78; DB 1; Length 390;  
 Best Local Similarity 31.3%; Pred. No. 1.53e+00;  
 Matches 10; Conservative 9; Mismatches 12; Indels 1; Gaps 1;  
 Db 157 LAIRAPHTYASHMKTRATRAVLGVWLVAF 188  
 QY 20 LQTRKPH-TGNHLKSGMTVMVKWLLMTF 50  
 RESULT 7  
 ID PER3\_RABIT STANDARD; PRT; 411 AA.  
 AC P46059;  
 DT 01-NOV-1995 (REL. 32, CREATED)  
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
 DE PROSTAGLANDIN E2 RECEPTOR, EP3 SUBTYPE (PROSTANOID EP3 RECEPTOR) (PGE  
 DE RECEPTOR, EP3 SUBTYPE)  
 OS PTGER3.  
 OS ORYCTOLAGUS CUNICULUS (RABBIT).  
 CC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 CC LAGOMORPHA; LEPORIDAE; ORYCTOLAGUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-NEW ZEALAND WHITE; TISSUE-KIDNEY CORTEX;  
 FA MEYER R.M., EMERSON R.B., TARNAG J.L., BREYER M.D., DAVIS L.S.,  
 RA ABROMSON R.M., FERRENBACH S.M.;  
 XT "Alternative splicing generates multiple isoforms of a rabbit  
 RT prostaglandin E2 receptor.";  
 FJ J. BIOL. CHEM. 269:6163-6169(1994).  
 CC -!- FUNCTION: RECEPTOR FOR PROSTAGLANDIN E2 (PGE2). THE ACTIVITY OF  
 THIS RECEPTOR CAN COUPLE TO BOTH THE INHIBITION OF ADENYLATE  
 CYCLASE MEDIATED BY G-I PROTEINS, AND TO AN ELEVATION OF  
 INTRACELLULAR CALCIUM. THE VARIOUS FORMS CAN INTERACT WITH  
 DIFFERENT SECOND MESSENGER SYSTEMS (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -!- TISSUE SPECIFICITY: IN THE KIDNEY, HIGH LEVELS OF EXPRESSION ARE  
 SEEN IN MEDULLARY THICK ASCENDING LIMB, AND LOWER LEVELS IN THE  
 CORTICAL AND MEDULLARY COLLECTING DUCTS.  
 CC -!- ALTERNATIVE PRODUCTS: AT LEAST FOUR FORMS THAT DIFFERS AT THE C-  
 TERMINUS ARE PRODUCED BY ALTERNATIVE SPLICING OF THIS RECEPTOR  
 GENE. THEY HAVE IDENTICAL LIGAND BINDING PROPERTIES BUT DIFFERENT  
 COUPLING PROPERTIES WITH G PROTEINS. THE SEQUENCE OF CLONE 77A IS  
 SHOWN:  
 CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib.ch](mailto:license@sib.ch)).  
 CC



CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

CC EMBL: M13466; G142578; -

DR PIR; A29836; A29836.

DR HYDROLASE; GLYCOSIDASE; MULTIGENE FAMILY.

KW SIMILAR 94 100 TO LACZ FROM E. COLI (AA 457-463).

FT ACT\_SITE 148 148 PROTON DONOR (POTENTIAL).

FT ACT\_SITE 303 303 NUCLEOPHILE (POTENTIAL).

SQ SEQUENCE 672 AA; 78053 MW; 1C5D0C23 CRC32;

Query Match 16.6%; Score 78; DB 1; Length 672;

Best Local Similarity 31.4%; Pred. No. 1.53e+00;

Matches 11; Conservative 12; Mismatches 10; Indels 2; Gaps 2;

Db 325 LMSYATARGADGIMFFQWRQSRAGAE-KFPGAMV 358

QY 5 LWGHLGVVK-ANGLLILQTRKPHNHLKTSGGMV 38

RESULT 10

ID AMSJ\_ERWAM STANDARD; PRT; 415 AA.

AC Q46637;

DT 01-NOV-1997 (REL. 35, CREATED)

DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)

DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

DE AMYLOVRAN BIOSYNTHESIS PROTEIN AMSJ.

GN AMSJ.

OS ERWINIA AMYLOVORA.

OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;

OC ERWINIA.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 95319333.

RA BUGERT P., GEIDER K.;

RT "Molecular analysis of the *ams* operon required for exopolysaccharide synthesis of *Erwinia amylovora*.";

RL MOL. MICROBIOL. 15:917-933(1995).

CC -!- FUNCTION: INVOLVED IN THE BIOSYNTHESIS OF AMYLOVRAN WHICH FUNCTIONS AS A VIRULENCE FACTOR.

CC -!- PATHWAY: EXOPOLYSACCHARIDE BIOSYNTHESIS.

CC -!- SIMILARITY: TO E. COLI COLANIC ACID BIOSYNTHESIS PROTEIN WCAK.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X77921; G600436; -

KW EXOPOLYSACCHARIDE SYNTHESIS.

SQ SEQUENCE 415 AA; 46397 MW; 85DCA577 CRC32;

Query Match 16.1%; Score 76; DB 1; Length 415;

Best Local Similarity 40.0%; Pred. No. 3.07e+00;

Matches 10; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

Db 64 KKSNSLVAKVKRLMPKIMAHIRG 88

QY 32 KTSGGMVTWKRWLLMTFMAGCRG 56

RESULT 11

ID AGGL\_RICCO STANDARD; PRT; 564 AA.

AC P06750;

DT 01-JAN-1988 (REL. 06, CREATED)

DT 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)

DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)

DE AGGLUTININ PRECURSOR (RRNA N-GLYCOSIDASE) (EC 3.2.2.22) (RCA).

OS RIKUNYOTA COMMUNIS (CASTOR BEAN).

OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EUDICOTYLEDONS; ROSIDAE; EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; TRACHEOPHYTA;

CC EUPHORBIALES; EUPHORBACEAE; RICINUS.

CC [1]

RN SEQUENCE FROM N.A.

RX MEDLINE; 86059449.

RA ROBERTS I.M., LAMB F.I., PAPPIN D.J.C., LORD J.M.;

RT "The primary sequence of Ricinus communis agglutinin. Comparison with ricin."

RL J. BIOL. CHEM. 260:15682-15686(1985).

RN RN

PP SEQUENCE OF 303-564.

RA ARAKI Y., YOSHIOKA Y., FUNATSU G.;

RT "The complete amino acid sequence of the B-chain of the Ricinus communis agglutinin isolated from large-grain castor bean seeds."

RL BIOCHIM. BIOPHYS. ACTA 872:277-285(1986).

RN RN

RP SEQUENCE OF 303-337.

RX MEDLINE; 80178723.

RA LIN T.-S., LI S.-L.;

RT "Purification and physicochemical properties of ricins and agglutinins from Ricinus communis."

RL EUR. J. BIOCHEM. 105:453-459(1980).

CC -!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF THE N-GLYCOSIDIC BOND AT ONE SPECIFIC ADENOSINE ON THE 28S RRNA.

CC -!- SIMILARITY: TO OTHER BACTERIAL AND PLANTS RIBOSOME-INACTIVATING PROTEINS. BELONGS TO TYPE 2 RIP.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: M12089; G169701; -

DR PIR; A24261; RLCSAG.

DR PROSITE; PS00275; SHIGA\_RICIN; 1.

DR PFAM; PF00161; RIP; 1.

DR PFAM; PF00652; Ricin\_B\_lectin; 2.

DR HSP; P02879; IAPG.

KW HYDROLASE; GLYCOSIDASE; TOXIN; DUPLICATION; REPEAT; GLYCOPROTEIN; LECTIN; SIGNAL.

FT SIGNAL 1.

FT CHAIN 24

FT PEPTIDE 25 290 AGGLUTININ A CHAIN.

FT CHAIN 303 302 LINKER PEPTIDE.

FT CHAIN 303 564 AGGLUTININ B CHAIN.

FT ACT\_SITE 200 200 BY SIMILARITY.

FT CARBOHYD 34 34 POTENTIAL.

FT CARBOHYD 259 259 POTENTIAL.

FT CARBOHYD 397 397

FT CARBOHYD 437 437 BY SIMILARITY.

FT DISULFID 282 306 BY SIMILARITY.

FT DISULFID 322 341 BY SIMILARITY.

FT DISULFID 365 382 BY SIMILARITY.

FT DISULFID 453 466 BY SIMILARITY.

FT DISULFID 492 509 BY SIMILARITY.

FT SIMILAR 451 564 31% IDENTITY WITH BETA-1,3-GLUCANASE FROM O-XANTHINEOLYTICA (472-548).

FT CONFLICT 331 331 F -> T (IN REF. 2).

FT CONFLICT 362 362 N -> D (IN REF. 2).

FT CONFLICT 374 374 R -> G (IN REF. 2).

FT CONFLICT 404 404 R -> T (IN REF. 2).

FT CONFLICT 552 552 F -> V (IN REF. 2).

SQ SEQUENCE 564 AA; 62851 MW; 72B3362E CRC32;

Query Match 16.1%; Score 76; DB 1; Length 564;

Best Local Similarity 17.1%; Pred. No. 3.07e+00;

Matches 7; Conservative 14; Mismatches 20; Indels 0; Gaps 0;

Db 394 LDNRRTINPRSGLVLAATSGNSGKTIVQNTNIVYSGWL 434

QY 5 LWGHLGVVKANGLLILQTRKPHNHLKTSGGMVTWKWL 45

```

RESULT 12
ID PRIA_RHURU STANDARD; PRT; 811 AA.
AC P05445;
DT 01-NOV-1988 (REL. 09, CREATED)
DT 01-NOV-1988 (REL. 09, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE PRIMOSOMAL PROTEIN N' (REPLICATION FACTOR Y).
GN PRIA.
OS RHODOSPIRILLUM RUBRUM.
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHODOSPIRILLACEAE;
OC RHODOSPIRILLUM.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 85251588.
RA FALK G., HAMPE A., WALKER J.E.;
RL Nucleotide sequence of the Rhodospirillum rubrum atp operon.;
RL BIOCHEM. J. 228:391-407(1985).
CC -1- FUNCTION: PRIA RECOGNIZES A SPECIFIC HAIRPIN SEQUENCE ON PHIX
CC SSDNA. THIS STRUCTURE IS THEN RECOGNIZED AND BOUND BY PROTEINS
CC PRIB AND PRIC. FORMATION OF THE PRIMOSOME PROCEEDS WITH THE
CC SUBSEQUENT ACTIONS OF DNAB, DNAC, DNAT AND PRIMASE. PRIA THEN
CC FUNCTIONS AS A HELICASE WITHIN THE PRIMOSOME (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE PRIA SUBFAMILY OF HELICASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; X02499; G46363;
DR EMBL; X02499; G46364; ALT_INIT.
DR PIR; S08579; S08579.
KW DNA REPLICATION; DNA-BINDING; ATP-BINDING; HELICASE; PRIMOSOME;
KW ZINC-FINGER.
FT NP_BIND 253 260 ATP (BY SIMILARITY).
FT SITE 349 352 DEER BOX.
FT 2N_FING 520 532 C4-TYPE (POTENTIAL).
FT 2N_FING 547 563 C4-TYPE (POTENTIAL).
SQ SEQUENCE 811 AA; 85574 MW; 5F50C3D8 CRC32;

Query Match 16.1%; Score 76; DB 1; Length 811;
Best Local Similarity 34.7%; Pred. No. 3.07e+00;
Matches 17; Conservative 12; Mismatches 16; Indels 4; Gaps 4;

Db 754 LGVPAP-LAMLRGRHR-LLKKAARGVKQVVRWLSLVSPGK 800
QY 9 LGVYKANGLLILQTRKPHNLKTS-GGMV-TMVKWLLMTMAGCR 55

RESULT 13
ID YSC1_THREL STANDARD; PRT; 93 AA.
AC P25124;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DT 01-MAY-1992 (REL. 22, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL PROTEIN IN SC5B 5' REGION (ORF) (FRAGMENT).
OS THERMUS AQUATICUS (SUBSP. FLAVUS).
OC BACTERIA; THERMUS/DEINOCOCCUS GROUP; THERMUS GROUP; THERMUS.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=AT-62.
RC MEDLINE; 91238680.
RA NISHIYAMA M., HORINOCHI S., BEPPU T.;
RT Characterization of an operon encoding succinyl-CoA synthetase and
RT malate dehydrogenase from Thermus flavus AT-62 and its expression in
RT Escherichia coli.;
RT MOL. GEN. GENET. 226:1-9(1991).
RL
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; X54073; G48175;
DR PIR; S15948; S15948.
KW HYPOTHETICAL PROTEIN.
FT NON_TER 1
SQ SEQUENCE 93 AA; 10011 MW; 0B029A19 CRC32;

Query Match 15.9%; Score 75; DB 1; Length 93;
Best Local Similarity 35.9%; Pred. No. 4.32e+00;
Matches 14; Conservative 8; Mismatches 14; Indels 3; Gaps 3;

Db 20 ALRPOLOVDTKLILKEGKEPHOGVYLE-DGSMVVV 57
QY 4 SLWHLGVVKNGLLIL-QTRKPHYG-NHLKTSGGNVTM 40

RESULT 14
ID UL04_HCMVA STANDARD; PRT; 152 AA.
AC P17146;
DT 01-AUG-1990 (REL. 15, CREATED)
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE EARLY GLYCOPROTEIN GP48 PRECURSOR.
GN UL4.
OS HUMAN CYTOMEGALOVIRUS (STRAIN AD169).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC BETAHERPESVIRINAE; CYTOMEGALOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90269039.
RA CHEE M.S., BANKIER A.T., BECK S., BOHNI R., BROWN C.M., CERNY R.,
RA HORSNELL T., HUTCHISON C.A. III, KOUZARIDES T., MARTIGNETTI J.A.,
RA PREDDIE E., SATCHWELL S.C., TOMLINSON P., WESTON K.M., BARRELL B.G.;
RT Analysis of the protein-coding content of the sequence of human
RT cytomegalovirus strain AD169.;
RL CUER. TOP. MICROBIOL. IMMUNOL. 154:125-169(1990).
CC -1- FUNCTION: SINCE GP48 WAS FOUND IN THE VIRION, IT IS CONSIDERED
CC AN EARLY STRUCTURAL PROTEIN.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND (POTENTIAL).
CC -1- PTM: N-GLYCOSYLATED AND POSSIBLY O-GLYCOSYLATED.
CC -1- SIMILARITY: BELONGS TO RL11 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; X17403; G59609;
DR PIR; S09767; VGBEY9.
KW EARLY PROTEIN; GLYCOPROTEIN; MEMBRANE; SIGNAL.
FT SIGNAL 1 25
FT CHAIN 26 152
FT CARBOHYD 48 48 EARLY GLYCOPROTEIN GP48.
FT CARBOHYD 53 53 POTENTIAL.
FT CARBOHYD 61 61 POTENTIAL.
FT CARBOHYD 69 69 POTENTIAL.
FT CARBOHYD 108 108 POTENTIAL.
FT CARBOHYD 112 112 POTENTIAL.
FT CARBOHYD 122 122 POTENTIAL.
FT CARBOHYD 139 139 POTENTIAL.
FT CARBOHYD 148 148 POTENTIAL.
SQ SEQUENCE 152 AA; 17750 MW; 0BDE403 CRC32;

Query Match 15.9%; Score 75; DB 1; Length 152;

```





\*\*\*\*\*  
MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Wed Dec 22 10:26:52 1999; MasPar time 10.53 Seconds  
Tabular output not generated. 305.877 Million cell updates/sec  
Title: >US-09-209-961-20  
Description: (1-59) from US09209961.pep  
Perfect Score: 471  
Sequence: 1 MYTSLWHLGVVKGANGLLIL.....MVKWLLMTFMAGCRGMII 59  
Scoring table: PAM 150  
Gap 11  
Searched: 179066 seqs, 54579741 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: sptrembl9  
1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phase 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus  
Statistics: Mean 36.449; Variance 61.032; scale 0.597  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.  
SUMMARIES  
Result No. Score Query Match Length DB ID Description Pred. No.  
1 471 100.0 59 14 O92288 P6.6 3.89e-86  
2 95 20.2 593 2 Q49695 B1496.F1.14. 6.74e-03  
3 94 20.0 591 2 O53150 HYPOTHETICAL 62.7 KD P 9.89e-03  
4 89 18.9 147 5 Q17666 COSMID C05E11. 6.49e-02  
5 88 18.7 607 2 P96589 YDAA PROTEIN. 9.40e-02  
6 83 17.6 387 6 O46657 PROTAGLANDIN EP3 RECE 5.73e-01  
7 82 17.4 129 5 Q95034 SUPEROXIDE DISMUTASE ( 8.16e-01  
8 82 17.4 198 5 Q27740 SUPEROXIDE DISMUTASE ( 8.16e-01  
9 80 17.0 101 14 Q88146 VPR PROTEIN. 1.64e+00  
10 80 17.0 373 1 Q26344 SENSORY TRANSDUCTION H 1.64e+00  
11 80 17.0 691 2 O54098 PUTATIVE MEMBRANE PROT 1.64e+00  
12 80 17.0 885 2 P74259 HYPOTHETICAL 101.5 KD 1.64e+00  
13 79 16.8 60 8 Q36334 UNIQUE ORF. 2.31e+00  
14 79 16.8 578 5 O17444 VESICULAR ACETYLCHOLIN 2.31e+00  
15 78 16.6 286 2 Q46525 FIMP. 3.26e+00  
16 78 16.6 393 4 O00326 PROTAGLANDIN EP3 RECE 3.26e+00  
17 78 16.6 394 4 O15192 EP3-III PROTEIN. 3.26e+00  
18 78 16.6 402 4 O00325 PROTAGLANDIN EP3 RECE 3.26e+00  
19 78 16.6 403 4 O15193 EP3-IV PROTEIN. 3.26e+00  
20 78 16.6 419 4 O15549 EP3-I PROTEIN. 3.26e+00

\*\*\*\*\*  
Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd  
\*\*\*\*\*

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Wed Dec 22 10:26:52 1999; MasPar time 10.53 Seconds  
Tabular output not generated. 305.877 Million cell updates/sec

Title: >US-09-209-961-20  
Description: (1-59) from US09209961.pep  
Perfect Score: 471  
Sequence: 1 MYTSLWHLGVVKGANGLLIL.....MVKWLLMTFMAGCRGMII 59

Scoring table: PAM 150  
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: sptrembl9  
1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phase 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 36.449; Variance 61.032; scale 0.597

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	471	100.0	59	14	O92288 P6.6.	3.89e-86
2	95	20.2	593	2	Q49695 B1496.F1.14.	6.74e-03
3	94	20.0	591	2	O53150 HYPOTHETICAL 62.7 KD P	9.89e-03
4	89	18.9	147	5	Q17666 COSMID C05E11.	6.49e-02
5	88	18.7	607	2	P96589 YDAA PROTEIN.	9.40e-02
6	83	17.6	387	6	O46657 PROTAGLANDIN EP3 RECE	5.73e-01
7	82	17.4	129	5	Q95034 SUPEROXIDE DISMUTASE (	8.16e-01
8	82	17.4	198	5	Q27740 SUPEROXIDE DISMUTASE (	8.16e-01
9	80	17.0	101	14	Q88146 VPR PROTEIN.	1.64e+00
10	80	17.0	373	1	O26344 SENSORY TRANSDUCTION H	1.64e+00
11	80	17.0	691	2	O54098 PUTATIVE MEMBRANE PROT	1.64e+00
12	80	17.0	885	2	P74259 HYPOTHETICAL 101.5 KD	1.64e+00
13	79	16.8	60	8	Q36334 UNIQUE ORF.	2.31e+00
14	79	16.8	578	5	O17444 VESICULAR ACETYLCHOLIN	2.31e+00
15	78	16.6	286	2	Q46525 FIMP.	3.26e+00
16	78	16.6	393	4	O00326 PROTAGLANDIN EP3 RECE	3.26e+00
17	78	16.6	394	4	O15192 EP3-III PROTEIN.	3.26e+00
18	78	16.6	402	4	O00325 PROTAGLANDIN EP3 RECE	3.26e+00
19	78	16.6	403	4	O15193 EP3-IV PROTEIN.	3.26e+00
20	78	16.6	419	4	O15549 EP3-I PROTEIN.	3.26e+00

21	78	16.6	422	4	O15190	EP3E PROTEIN.	3.26e+00
22	78	16.6	422	4	O15189	EP3-VI PROTEIN.	3.26e+00
23	78	16.6	431	4	O15188	EP3-V PROTEIN.	3.26e+00
24	78	16.6	445	8	O79555	NADH DEHYDROGENASE SUB	3.26e+00
25	78	16.6	454	4	O15191	EP3F PROTEIN.	3.26e+00
26	78	16.6	535	3	Q01165	TRANSPOSASE.	3.26e+00
27	78	16.6	692	2	Q45772	OUTER MEMBRANE PROTEIN	3.26e+00
28	78	16.6	792	2	O83999	CATION-TRANSPORTING AT	3.26e+00
29	77	16.3	142	2	Q44529	POTENTIAL MOLYBDENUM-P	4.57e+00
30	77	16.3	208	8	O21609	NADH DEHYDROGENASE SUB	4.57e+00
31	77	16.3	311	2	O65393	NAC GENE.	4.57e+00
32	77	16.3	359	5	O15839	L4171.5.	4.57e+00
33	77	16.3	434	1	O57980	434AA LONG HYPOTHETICA	4.57e+00
34	77	16.3	444	3	Q00921	MOX METHANOL OXIDASE.	4.57e+00
35	77	16.3	473	2	O24723	HYPOTHETICAL 49.1 KD P	4.57e+00
36	77	16.3	1073	5	O21885	COSMID R09H3.	4.57e+00
37	76	16.1	173	8	O79712	NADH DEHYDROGENASE SUB	6.39e+00
38	76	16.1	224	8	Q95896	NADH DEHYDROGENASE SUB	6.39e+00
39	76	16.1	328	3	O59938	FAMILY F XLANASE.	6.39e+00
40	76	16.1	448	10	O23100	BAC TM018A10.	6.39e+00
41	76	16.1	595	13	Q91751	FURIN (FRAGMENT).	6.39e+00
42	76	16.1	609	2	O84866	PROTEASE CONTAINING IR	6.39e+00
43	76	16.1	690	2	O33377	TRANSFERRIN-BINDING PR	6.39e+00
44	76	16.1	743	10	O82302	T32F12.27 PROTEIN.	6.39e+00
45	76	16.1	1007	6	O46432	LYSOSOMAL ALPHA-MANNOS	6.39e+00

ALIGNMENTS

RESULT 1	PRELIMINARY;	PRT;	59 AA.
ID O92288;			
AC O92288;			
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)			
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)			
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)			
DE P6.6.			
CS PORCINE CIRCOVIRUS.			
OC VIRUSES; SSDNA VIRUSES; CIRCOVIRIDAE; CIRCOVIRUS.			
RN [1]			
RP SEQUENCE FROM N.A.			
RC STRAIN-412;			
RA WANG L., WILLSON P., CHOW B., GIBBONS E., BABIUK L.;			
RT "Emergence of a new porcine circovirus.;"			
RL SUBMITTED (AUG-1998) TO EMBL/GENBANK/DBJ DATA BANKS.			
DR ENBL; AF085695; G3668364;			
SQ SEQUENCE 59 AA; 6593 MW; A40B778A CRC32;			
Query Match 100.0%; Score 471; DB 14; Length 59;			
Best Local Similarity 100.0%; Pred. No. 3.89e-86;			
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			

DB 1	MYTSLWHLGVVKGANGLLILQTRPHTGNHLKTSGGVMTYVKWLLMTFMAGCRGMII 59	
QY 1	MYTSLWHLGVVKGANGLLILQTRPHTGNHLKTSGGVMTYVKWLLMTFMAGCRGMII 59	

RESULT 2	PRELIMINARY;	PRT;	593 AA.
ID Q49695;			
AC Q49695;			
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)			
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)			
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)			
DE B1496.F1.14.			
CS MYCOBACTERIUM LEPRAE.			
OC MYCOBACTERIA; FRMTCUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;			
OC ACTINOMYCETALES; CORYNEBACTERIACEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.			
RN [1]			
RP SEQUENCE FROM N.A.			
RC ROBINSON K.;			
RL SUBMITTED (NOV-1993) TO EMBL/GENBANK/DBJ DATA BANKS.			
K/ [2]			
NP SEQUENCE FROM N.A.			



[illegible]

CC FAMILY.  
 DR EMBL; 249819; E183284; --  
 DR PROSITE; PS00088; SOD\_MN; 1.  
 DR PFAM; PF00081; sodfe; 1.  
 KW OXIDOREDUCTASE.  
 SQ SEQUENCE 198 AA; 22734 MW; 93C9D2D2 CRC32;  
 Query Match 17.4%; Score 82; DB 5; Length 198;  
 Best Local Similarity 34.1%; Pred. No. 8.16e+01;  
 Matches 15; Conservative 11; Mismatches 15; Indels 3; Gaps 2;  
 Db 119 FGCGWGLANNKNNKLVLIQTH--DAGNPIKDNIGIPILTCDIW 160  
 QY 2 YTLWGLGVVKKANGLLILQTRKPHGTHLKTSGGMVTWV-KKW 44  
 RESULT 9  
 ID Q88146 PRELIMINARY; PRT; 101 AA.  
 AC Q88146;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE VPR PROTEIN.  
 GN VPR.  
 OS SIMIAN IMMUNODEFICIENCY VIRUS (SIV-AGM).  
 OC VIRUSES; RETROID VIRUSES; RETROVIRIDAE; LENTIVIRUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 92124755.  
 RA NOVEMBRE F.J., HIRSCH V.M., MCCLURE H.M., FULTZ P.N., JOHNSON P.R.;  
 RT "SIV from stump-tailed macaques: molecular characterization of a  
 highly transmissible primate lentivirus.";  
 RL VIROLOGY 186:783-787(1992).  
 DR EMBL; M83293; G334805;  
 DR PFAM; PF00522; VPR; 1.  
 SQ SEQUENCE 101 AA; 11448 MW; BB3B6F14 CRC32;  
 Query Match 17.0%; Score 80; DB 14; Length 101;  
 Best Local Similarity 32.1%; Pred. No. 1.64e+00;  
 Matches 9; Conservative 9; Mismatches 9; Indels 1; Gaps 1;  
 Db 52 GDTLEGAGELIKILQIR-ALFPHFRGGCR 78  
 QY 28 GNHLKTSGGMTVMVKKWLLMTFMAGCR 55  
 RESULT 10  
 ID Q26544 PRELIMINARY; PRT; 373 AA.  
 AC Q26544;  
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE SENSORY TRANSDUCTION HISTIDINE KINASE.  
 GN MTH444.  
 OS METHANOBACTERIUM THERMOAUTOTROPHICUM.  
 OC ARCHAEA; EURYARCHAEOTA; METHANOBACTERIALES; METHANOBACTERIACEAE;  
 OC METHANOBACTERIUM.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-DELTA H;  
 RX MEDLINE; 98037514.  
 RA SMITH D.R., DOUCETTE-STAMM L.A., DELOUGHERY C., LEE H.-M., DUBOIS J.,  
 RA ALDRIDGE T., BASHIRZADEH R., BLAKELY D., COOK R., GILBERT K.,  
 RA HARRISON D., HOANG L., KEAGLE P., LUM W., POTHIER B., QIU D.,  
 RA SPADAFORA R., VICARE R., WANG Y., WIERZBORSKI J., GIBSON R.,  
 RA JIWANI N., CARUSO A., BUSH D., SAFER H., FATWELL D., PRABHAKAR S.,  
 RA MCDUGALL S., SHIMER G., GOYAL A., PIETROVSKI S., CHURCH G.M.,  
 RA DANIELS C.J., MAO J.-I., RICE P., NOLLING J., REEVE J.N.;  
 RT "Complete genome sequence of Methanobacterium thermoautotrophicum  
 deltaH: functional analysis and comparative genomics.";  
 RL J. BACTERIOLOGY 179:7135-7155(1997).  
 DR EMBL; AE000828; G2621510;  
 DR PFAM; PF00512; signal; 1.

SQ SEQUENCE 373 AA; 42263 MW; A636962D CRC32;  
 Query Match 17.0%; Score 80; DB 1; Length 373;  
 Best Local Similarity 37.5%; Pred. No. 1.64e+00;  
 Matches 18; Conservative 6; Mismatches 20; Indels 4; Gaps 2;  
 Db 76 WGVLVAVFVALLIQLQPMTRAVG--LNISDELFLRLV--LLLTFFMVA 119  
 QY 6 WGH-GVVKANGLLILQTRKPHGTHLKTSGGMVTWVKKWLLMTFMAG 53  
 RESULT 11  
 ID OS4098 PRELIMINARY; PRT; 691 AA.  
 AC OS4098;  
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE PUTATIVE MEMBRANE PROTEIN.  
 GN SC10A5.05.  
 OS STREPTOMYCES COELICOLOR.  
 OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIACEAE;  
 CC ACTINOMYCETALES; STREPTOMYCINAE; STREPTOMYCETACEAE; STREPTOMYCES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA MURPHY L., HARRIS D.;  
 RL SUBMITTED (JAN-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;  
 RL SUBMITTED (JAN-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA REDENBACH M., KIESER H.M., DENAPATE D., EICHNER A., CULLUM J.,  
 RA KINASHI H., HOPWOOD D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
 RL MOL. MICROBIOL. 21:77-96(1996).  
 DR EMBL; AL021529; E1247672;  
 SQ SEQUENCE 691 AA; 75164 MW; C92968B4 CRC32;  
 Query Match 17.0%; Score 80; DB 2; Length 691;  
 Best Local Similarity 29.5%; Pred. No. 1.64e+00;  
 Matches 13; Conservative 13; Mismatches 16; Indels 2; Gaps 2;  
 Db 195 GLVLDDTNRAPTADYEIKPEHGGLAGFALIFLLRAFGSGCAAL 238  
 QY 16 GLILQTRKPHGTHN-LKTS-GGMVTWVKKWLLMTFMAGCRGM 57  
 RESULT 12  
 ID P74259 PRELIMINARY; PRT; 885 AA.  
 AC P74259;  
 DT 01-FEB-1997 (TREMBLREL. 02, CREATED)  
 DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)  
 DE HYPOTHETICAL 101.5 KD PROTEIN.  
 OS SYNECHOCYSTIS SP. (STRAIN PCC 6803).  
 OC BACTERIA; CYANOBACTERIA; CHROCOCCALES; SYNECHOCYSTIS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-PC6803;  
 RA TABATA S.;  
 RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-PC6803;  
 RA MEDLINE; 97061201  
 RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,  
 RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,

RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,  
 RA SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,  
 RA TABATA S.;  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 RT Synechocystis sp. PCC6803. II. Sequence determination of the entire  
 RT genome and assignment of potential protein-coding regions.";  
 RL DNA RES. 3:109-136(1996).  
 DR EMBL; D90913; D1019086;  
 DR PFAM; PF00128; alpha-amylase; 1.  
 KW HYPOTHETICAL PROTEIN.  
 SQ SEQUENCE 885 AA; 101491 MW; C9C80BAE CRC32;

Query Match 17.0%; Score 80; DB 2; Length 885;  
 Best Local Similarity 39.5%; Pred. No. 1.64e+00;  
 Matches 15; Conservative 7; Mismatches 13; Indels 3; Gaps 3;  
 Db 409 MYGODLNHOLPIVRAI-LLEMORRINTGADGIRVDGG 445  
 QY 1 MYT-SLWGHGVVRANGLLIQLTRPHTGNH-LKTSGG 36

RESULT 13  
 ID Q36334 PRELIMINARY; PRT; 60 AA.  
 AC Q36334;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE UNIQUE ORF.  
 GN PUTATIVE ORF60.  
 OS CHONDRUS CRISPUS (CARRAGHEEN).  
 OC MITOCHONDRION.  
 OC EUKARYOTA; RHODOPHYTA; FLORIDOPHYCEAE; GIGARTINALES; GIGARTINACEAE;  
 OC CHONDRUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-APICES;  
 RX MEDLINE; 95341681.  
 RA LEBLANC C., BOYEN C., RICHARD O., BONNARD G., GRIENENBERGER J.M.,  
 RA KLOAREG B.;  
 RT "Complete sequence of the mitochondrial DNA of the rhodophyte  
 RT Chondrus crispus (gigartinales). Gene content and genome  
 RT organization.";  
 RL J. MOL. BIOL. 250:484-495(1995).  
 DR EMBL; Z47547; G1019071;  
 KW MITOCHONDRION.  
 SQ SEQUENCE 60 AA; 6633 MW; 49A006C5 CRC32;

Query Match 16.8%; Score 79; DB 8; Length 60;  
 Best Local Similarity 42.9%; Pred. No. 2.31e+00;  
 Matches 6; Conservative 7; Mismatches 1; Indels 0; Gaps 0;  
 Db 1 MYRHWLLMLAFKS 14  
 QY 40 MYKKWLLMTFMAG 53

RESULT 14  
 ID O17444 PRELIMINARY; PRT; 578 AA.  
 AC O17444;  
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE VESICULAR ACETYLCHOLINE TRANSPORTER.  
 GN VACHT.  
 OS DROSOPHILA MELANOGASTER (FRUIT FLY).  
 OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;  
 OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;  
 OC DROSOPHILIDAE; DROSOPHILA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA KITAMOTO T., WANG W., SALVATERRA P.M.;  
 RL J. BIOL. CHEM. 0:0-0(1997).  
 DR EMBL; AF030197; G2625056;

SQ SEQUENCE 578 AA; 64345 MW; 4AE4211D CRC32;  
 Query Match 16.8%; Score 79; DB 5; Length 578;  
 Best Local Similarity 31.8%; Pred. No. 2.31e+00;  
 Matches 14; Conservative 9; Mismatches 18; Indels 3; Gaps 3;  
 Db 22H DGLMLLLVMKPKVEA-MKOSKDVQVQVIFIRLLMDPYIAVCAG 270  
 QY 15 NGLLIQLTRPHTGNHKTSGMVT-MVKKWLILLM-TFMAGCRG 56

RESULT 15  
 ID Q46525 PRELIMINARY; PRT; 286 AA.  
 AC Q46525;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE FIMP.  
 GN FIMP.  
 OS BACTERIOIDES NODOSUS (DICHELOBACTER NODOSUS).  
 OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; CARDIOBACTERIACEAE;  
 OC DICHELOBACTER.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A198;  
 RX MEDLINE; 95369727.  
 RA JOHNSTON J.L., BILLINGTON S.J., HARING V., ROOD J.I.;  
 RT "Identification of fimbrial assembly genes from Dichelobacter  
 RT nodosus: evidence that fimp encodes the type-IV prepilin peptidase.";  
 RL GENE 161:21-26(1995).  
 DR EMBL; U17138; G790211;  
 SQ SEQUENCE 286 AA; 31863 MW; EA572E0E CRC32;

Query Match 16.6%; Score 78; DB 2; Length 286;  
 Best Local Similarity 28.8%; Pred. No. 3.26e+00;  
 Matches 12; Conservative 14; Mismatches 14; Indels 2; Gaps 2;  
 Db 67 LLVPASRCPHCGRIRAIEN-IPVIS-WLFLKGCSCGGAII 106  
 QY 17 LLILQTRPHTGNHKTSGMVTMYKKWLLMTFMAGCRGMI 58

Search completed: Wed Dec 22 10:27:07 1999  
 Job time : 15 secs.

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Run on: Wed Dec 22 10:28:27 1999; MasPar time 12.06 Seconds  
93.495 Million cell updates/sec  
Tabular output not generated.

```
>US-09-209-961-21
Description: (1-53) from US09209961.pep
Perfect Score: 416
Sequence: 1 MVEIILHGFKGWGFKIKFSE.....VETTFERSAEAYVVHISRGL 53
```

Scoring table: PAM 150  
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database:

a-genesq35

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39

```
Statistics:      Mean 26.290;  Variance 115.045;  scale 0.229
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	81	19.5	317	28	W33721	Human melanocortin-1	3.82e+01
2	81	19.5	317	36	W79884	Melanocortin-1 recept	3.82e+01
3	81	19.5	317	22	W19706	Melanocortin-1 recept	3.82e+01
4	81	19.5	317	9	R49725	Sequence of a polypep	3.82e+01
5	81	18.5	317	8	R43571	Human MSH-R.	3.82e+01
6	80	19.2	317	31	W37627	Human melanocyte stim	4.53e+01
7	79	19.0	246	36	W77312	Polypeptide sequence	5.37e+01
8	79	19.0	404	31	W59035	Human G-protein coupl	5.37e+01
9	79	19.0	404	28	W31344	Human G-protein coupl	5.37e+01
10	79	19.0	404	28	W34984	Human G-protein coupl	5.37e+01
11	79	19.0	1099	38	W29683	Human Na-K-2Cl cotran	5.37e+01
12	78	18.8	197	34	W77749	Staphylococcus aureus	5.37e+01
13	78	18.8	400	13	R71932	I-branching enzyme.	6.36e+01
14	78	18.8	400	16	R92474	Beta-1,6-N-acetylgluc	6.36e+01
15	78	18.8	400	29	W47184	Human beta-1,6-N-acet	6.36e+01
16	78	18.8	400	31	W56522	Full length amino aci	6.36e+01

W79684;  
 17-DEC-1998 (first entry)  
 Melanocortin-1 receptor.  
 Human melanocortin-1 receptor; MC1; homology; transmembrane domain;  
 G-protein; MC3; therapeutic.  
 OS Homo sapiens.  
 PN US5817787-A.  
 PD 06-OCT-1998. 842045.  
 PF 23-APR-1997; US-200711.  
 PR 17-FEB-1994; US-672109.  
 PR 27-JUN-1996; US-672109.  
 PR 23-APR-1997; US-842045.  
 PA (UNMI) UNIV MICHIGAN.  
 PI Gantz I, Yamada T;  
 DR WPI; 98-556471/47.  
 DR N-PSDB; V62349.  
 PT DNA encoding melanocortin-5 receptor - useful in hybridisation  
 assays for melanocortin-5 receptor nucleic acids  
 PS Disclosure; Column 29-32; 58pp; English.  
 CC The present sequence represents the human melanocortin-1 (MC1) receptor.  
 CC the gene of which has been localised to chromosome 16q24.3. To obtain  
 CC the DNA sequence, RNA was first isolated from HeLa cells and U937 cells  
 CC and then subjected to RT-PCR, using primers based on the homologous  
 CC regions of the third and sixth transmembrane domains of G-protein linked  
 CC receptors. This sequence as well as MC3 (V62351), were then used to  
 CC construct oligonucleotides that could be used to search genomic DNA for  
 CC other members of the receptor family. These genes and their products may  
 CC be used to provide therapeutic vehicles for the treatment of processes  
 CC involving the function of melanocortin receptors.  
 SQ Sequence 317 AA;

Query Match 19.5%; Score 81; DB 36; Length 317;  
 Best Local Similarity 39.3%; Pred. No. 3.82e+01;  
 Matches 11; Conservative 5; Mismatches 11; Indels 1; Gaps 1;  
 Db 169 w-vasvstlfiaydhvavllclvfv 195  
 QY 11 WGVFKIFSELXIHGYTDIVLVVTFV 38

## RESULT

ID W19706 standard; Protein; 317 AA.  
 AC W19706;  
 DT 19-AUG-1997 (first entry)  
 DE Melanocortin-1 receptor.  
 KW Melanocortin-3; MC3R; MC1R; MC2R; MC4R; MC5R; human; mouse; melanocyte;  
 KW pro-opiomelanocortin; adrenal cortical function; behaviour; learning;  
 KW memory; cardiovascular system; analgesia; thermoregulation; prolactin;  
 KW neurohumoral agent; biogenic amine.  
 OS Homo sapiens.  
 PN US5622860-A.  
 PD 22-APR-1997.  
 PF 17-FEB-1994; 200711.  
 PR 17-FEB-1994; US-200711.  
 PA (UNMI) UNIV MICHIGAN.  
 PI Gantz I, Yamada T;  
 DR WPI; 97-244394/22.  
 DR N-PSDB; T68792.  
 PT Nucleic acid molecules encoding melanocortin receptors - useful to  
 transfect mammalian cells lacking endogenous receptors to induce  
 their expression  
 PS Example 1; Column 29-32; 58pp; English.  
 CC W19703-W19707 represent the human and mouse melanocortin (MC) receptors.  
 CC This sequence represents the MC1R, and the gene encoding it is located at  
 CC chromosome locus 16q24.3. MCs are products of pro-opiomelanocortin  
 CC post-translational processing, and are known to have a broad array of  
 CC physiological actions. MCs are known to have effects on adrenal cortical  
 CC functions and on melanocytes, as well as affecting behaviour, learning,  
 CC memory, control of the cardiovascular system, analgesia, thermoregulation  
 CC and the release of other neurohumoral agents (such as prolactin and  
 CC biogenic amines). The nucleic acids can be used to transfect mammalian  
 CC cells lacking endogenous MC receptors to induce their expression. These  
 CC sequences can also be used to screen and identify drugs which

CC specifically react with MCRs on the surface of a cell. The drugs can then  
 CC be used for treating diseases which have MCRs implicated as one of their  
 CC causes. Vectors containing these sequences can also be used to treat the  
 CC diseases.  
 SQ Sequence 317 AA;

Query Match 19.5%; Score 81; DB 22; Length 317;  
 Best Local Similarity 39.3%; Pred. No. 3.82e+01;  
 Matches 11; Conservative 5; Mismatches 11; Indels 1; Gaps 1;  
 Db 169 w-vasvstlfiaydhvavllclvfv 195  
 QY 11 WGVFKIFSELXIHGYTDIVLVVTFV 38

## RESULT

ID R49725 standard; Protein; 317 AA.  
 AC R49725;  
 DT 20-AUG-1994 (first entry)  
 DE Sequence of a polypeptide (PP) with the properties of  
 DE melanotropic hormone receptor  
 KW Melanocyte stimulating hormone receptor;  
 KW diagnosis; therapy.  
 OS Homo sapiens.  
 PN WO9404674-A.  
 PD 03-MAR-1994.  
 PF 20-AUG-1993; DK02773.  
 PR 21-AUG-1992; DK-001046.  
 PR 10-SEP-1992; DK-001118.  
 PR 05-MAY-1993; DK-000528.  
 PA (CHHA) CHHAJLANI V.  
 PA (WIKB) WIKBERG J.  
 PI Chhajlani V, Wikberg J;  
 DR WPI; 94-083193/10.  
 DR N-PSDB; Q44340.  
 PT Novel DNA fragments encoding novel peptides - having properties  
 PT of melanotropic hormone receptors esp. encoding melanocyte  
 PT stimulating hormone receptors, are used to diagnose e.g. skin  
 PT cancers or vitiligo  
 PS Claim 10; Page 95-97; 134pp; English.  
 CC The DNA fragment with the sequence in Q44340 is derived from a human  
 CC cDNA library. The encoded polypeptide (R49725) constitutes the  
 CC entire polypeptide of a melanocyte stimulating hormone (MSH)  
 CC receptor. The mRNA of this clone was found only in the melanoma  
 CC cells and not in other tissues examined including brain, thymus,  
 CC parathyroid gland, parotid gland, salivary gland, adrenal gland,  
 CC testis, liver, lung, heart, spleen, skeletal muscle, intestine and  
 CC colon. Transmembrane sites in the polypeptide corresp. to bps 286-  
 CC 351, 394-465, 517-588, 640-711, 733-804, 898-972, 997-1088.  
 CC Glycosylation sites are found at AA residues 15 and 29; possible  
 CC phosphorylation sites are found at AA residues 42-45, 151-154,  
 CC 306-308; and a possible palmitoylation site is found at AA  
 CC residue 316.  
 SQ Sequence 317 AA;

Query Match 19.5%; Score 81; DB 9; Length 317;  
 Best Local Similarity 39.3%; Pred. No. 3.82e+01;  
 Matches 11; Conservative 5; Mismatches 11; Indels 1; Gaps 1;  
 Db 169 w-vasvstlfiaydhvavllclvfv 195  
 QY 11 WGVFKIFSELXIHGYTDIVLVVTFV 38

## RESULT

ID R43571 standard; Protein; 317 AA.  
 AC R43571;  
 DT 22-MAY-1994 (first entry)  
 DE Human MSH-R.  
 KW Melanocyte stimulating hormone receptor; mouse; genetic disease;  
 KW isolation; screening; antagonists.  
 OS Homo sapiens.  
 PN WO9321316-A.

PD 28-OCT-1993. U03247.  
 PF 07-APR-1993; US-866979.  
 PR 10-APR-1992; US-866979.  
 PA (UYOR-) UNIV OREGON HEALTH SCI.  
 PI Cone RD, Mountjoy KG;  
 DR WPI: 93-351726/44.  
 DR N-PSDB; Q51226.  
 PT Melanocyte stimulating hormone receptors and corresp. DNA - from  
 PT humans and mice, used for diagnosis of genetic disease and for  
 PT screening agonists  
 PS Claim 8; Fig 1B; 46pp; English.  
 CC The sequence is that of the human melanocyte-stimulating hormone  
 CC receptor isolated from a human melanoma cDNA library using a pair  
 CC of degenerate PCR primers corresp. to the putative third and sixth  
 CC transmembrane regions of G-protein coupled receptors. The sequence  
 CC can be used as a probe for the detection and diagnosis of genetic  
 CC diseases, or for the isolation of novel mammalian receptors and genes.  
 CC It may also be used for screening and quantitating cpds. as inhibitors  
 CC of agonist binding to the receptor.  
 CC See also R43571.  
 SQ Sequence 317 AA;

Query Match 19.5%; Score 81; DB 8; Length 317;  
 Best Local Similarity 39.3%; Pred. No. 3.82e+01;  
 Matches 11; Conservative 5; Mismatches 11; Indels 1; Gaps 1;

Db 169 w-vasvfvstlfi-gyydhvavllclvfv 195  
 QY 11 WGVFKIRFSELYIHGYTDIVLVVFTVF 38

RESULT 6  
 ID W37827 standard; Protein; 317 AA.  
 AC W37827;  
 DT 28-JUL-1998 (first entry)  
 DE Human melanocyte stimulating hormone-1 receptor.  
 KW Human melanocyte stimulating hormone-1 receptor gene; MCRL1;  
 KW MCRL1; agonist; antagonist; feeding; metabolic disorder;  
 KW eating disorder; anorexia; obesity; cachexia; cancer; inhibition.  
 OS Homo sapiens.  
 PN W09810058-A2.  
 PD 12-MAR-1998.  
 PF 04-SEP-1997; U15565.  
 PR 04-SEP-1996; US-706281.  
 PA (UYOR-) UNIV OREGON HEALTH SCI.  
 PI Boston BA, Chen W, Cone RD, Fan W, Kesterton RA,  
 PI Lu D;  
 DR WPI: 98-193618/17.  
 DR N-PSDB; V19136.  
 PT Identifying melanocortin receptor agonists and antagonists - using a  
 PT panel of recombinant mammalian cells expressing alpha-melanocyte  
 PT stimulating hormone, ACTH, MC-3, MC-4 and MC-5 receptors  
 PS Example 2B; Fig 2A-B; 121pp; English.  
 CC This is the amino acid sequence of the human melanocyte stimulating  
 CC hormone receptor (MCRL1). The MCR agonists and antagonists can be used  
 CC for modifying feeding behaviour in an animal. The antagonists can be  
 CC used for stimulating feeding while the agonists can be used for  
 CC inhibiting feeding. They can be used for the treatment of eating  
 CC disorders such as anorexia and obesity, and other pathological weight  
 CC and eating-related disorders. They can also be used to treat failure  
 CC to thrive disorders and disease-related cachexia, such as occurs in  
 CC cancer patients, as well as other metabolic disorders.  
 SQ Sequence 317 AA;

Query Match 19.2%; Score 80; DB 31; Length 317;  
 Best Local Similarity 44.8%; Pred. No. 4.53e+01;  
 Matches 13; Conservative 5; Mismatches 8; Indels 3; Gaps 3;

Db 169 w-vasvfvstlfi-gyydhvavllclvfv 195  
 QY 11 WGVFKIRFSELYIHGYTD-IVLVVFTVF 38

RESULT 7  
 ID W77317 standard; Protein; 246 AA.  
 AC W77317;  
 DT 16-DEC-1998 (first entry)  
 DE Polypeptide sequence of p28 Bap31/CDM.  
 KW Apoptosis; p28 Bap31/CDM; modulate; p28 Bap31; human; p20 product;  
 KW pro-FUICE/Bcl-2 protein; Bcl-3; Bcl-XL; degenerative disease; neoplasia.  
 OS Homo sapiens.  
 FH Key.  
 FT Peptide  
 FT Peptide 2..11  
 FT Domain /note= "sequence homologous to human Bap31"  
 FT Domain 7..27  
 FT Domain /note= "predicted transmembrane domain TM1"  
 FT Domain 43..64  
 FT Domain /note= "predicted transmembrane domain TM2"  
 FT Domain 102..123  
 FT Domain /note= "predicted transmembrane domain TM3"  
 FT Peptide 161..165  
 FT Peptide /note= "caspase recognition site"  
 FT Cleavage\_site 164..165  
 FT Peptide 235..239  
 FT Peptide /note= "caspase recognition site"  
 FT Cleavage\_site 238..239  
 FT Peptide 243..246  
 FT Peptide /note= "ER retention signal sequence"  
 PN W09839434-A1.  
 PD 11-SEP-1998.  
 PR 02-MAR-1998; IB0706.  
 PR 03-MAR-1997; CA-198988.  
 PA (UYMC-) UNIV MCGILL.  
 PI Branton PE, Ng WHF, Nguyen M, Shore GC;  
 ER WPI: 98-503359/43.  
 PT Polypeptide-modulating apoptosis, p28 Bap31 - useful to diagnose  
 PT diseases involving altered apoptosis, e.g. degenerative diseases or  
 PT neoplasia and to detect compounds modulating apoptosis  
 PS Claim 7; Fig 2C; 130pp; English.  
 CC This represents the amino acid sequence of the p28 Bap31/CDM polypeptide  
 CC that can modulate apoptosis. The polypeptide p28 Bap31, fragments and  
 CC encoding nucleic acids are useful to generate or identify compounds which  
 CC alter the biological activity of p28 Bap31 and thus modulate apoptosis  
 CC when administered to cells (e.g. mammalian and especially human or rodent  
 CC cells). Such compounds may affect e.g. p28 Bap31 cleavage to produce p20  
 CC product, formation of complex with pro-FUICE/Bcl-2 protein (especially  
 CC Bcl-2 or Bcl-XL), specific binding to an antibody or p28 Bap31 expression  
 CC in the cell. The compounds are especially useful to inhibit apoptosis in  
 CC mammals with degenerative diseases or to increase apoptosis in mammals  
 CC with neoplasia. p28 Bap31 can also be used diagnostically, by measuring  
 CC levels in mammalian samples whereby a reduction is indicative of the  
 CC presence or likelihood of a disease caused by decreased apoptosis  
 CC (neoplasia) and an increase indicates the presence of a disease caused  
 CC by increased apoptosis (especially a degenerative disease). Antibodies  
 CC specifically binding p28 Bap31/p20 can similarly be used to diagnose  
 CC diseases/likelihood of diseases involving altered apoptosis.  
 SQ Sequence 246 AA;

Query Match 19.0%; Score 79; DB 36; Length 246;  
 Best Local Similarity 38.7%; Pred. No. 5.37e+01;  
 Matches 12; Conservative 3; Mismatches 8; Indels 2; Gaps 2;

Db 31 rwqkifsrllvvellysgntffvllvllv 61  
 QY 10 WGVFKIRFSELYIHGYTDIVLVVFTVF 38

RESULT 8  
 ID W59335 standard; Protein; 404 AA.  
 AC W59335;  
 DT 04-AUG-1998 (first entry)  
 DE Human G-protein coupled receptor protein.  
 KW G-protein coupled receptor; human; cerebellum; adrenomedullin; asthma;  
 KW screening; agonist; antagonist; diagnosis; disease; osteoporosis;  
 KW Parkinson's disease; acute heart failure; urinary retention.  
 OS Homo sapiens.



PN US5763218-A.  
 PD 09-JUN-1998.  
 PF 14-AUG-1996; 696770.  
 PR 20-MAY-1996; US-017954.  
 PR 14-AUG-1996; US-696770.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PA (TAKE-) TAKEDA.  
 PI Fujii R, Hinuma S, Li Y, Ruben SM, Soppet DR;  
 DR N-PSDB; V11726.  
 DR N-PSDB; V11726.  
 PT DNA encoding G-protein coupled receptor polypeptide - useful for  
 PT producing recombinant polypeptide, e.g. in diagnosis  
 PS Claim 1; Fig 1; 19pp; English.  
 CC This sequence represents a novel human cerebellum G-protein coupled  
 CC receptor that binds adrenomedullin. This protein can be used in screening  
 CC assays for agonists or antagonists of the receptor and for diagnosis of  
 CC diseases associated with abnormal expression of the receptor, such as  
 CC asthma, Parkinson's disease, acute heart failure, urinary retention and  
 CC osteoporosis.  
 CC Sequence 404 AA;

Query Match 19.0%; Score 79; DB 31; Length 404;  
 Best Local Similarity 34.4%; Pred. No. 5.37e+01;  
 Matches 11; Conservative 9; Mismatches 9; Indels 3; Gaps 3;

Db 121 wgsfscrthfyfvmysiffvlclsv-dr 151  
 QY 11 WGVFKIKFSE-LY-INGYTDIVLVVFTVFER 40

RESULT 9  
 ID W31344 standard; Protein; 404 AA.  
 AC W31344;  
 DE 22-MAY-1998 (first entry)  
 DE Human G-protein coupled receptor.  
 KW Human; G-protein coupled receptor; identification; activator;  
 KW inhibitor; diagnosis; treatment; asthma; Parkinson's disease;  
 KW acute heart failure; urinary retention; osteoporosis.  
 OS Homo sapiens.  
 PN WO9744360-A1.  
 PD 27-NOV-1997.  
 PF 14-AUG-1996; U13239.  
 PR 20-MAY-1996; WO-U07622.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PA (TAKE-) TAKEDA CHEM IND LTD.  
 PI Fujii R, Hinuma S, Li Y, Ruben SM, Soppet DR;  
 DR N-PSDB; T89983.  
 DR Human G-coupled receptor and related DNA - useful for identifying  
 PT (antagonists to treat, e.g. asthma, Parkinson's disease, acute  
 PT heart failure, urinary retention and osteoporosis  
 PS Claim 7; Pages 35-37; 51pp; English.  
 CC The present sequence is a human G-protein coupled receptor.  
 CC which can be used to identify activators and inhibitors. The  
 CC presence of the receptor in a sample can also be used in diagnosis  
 CC of a disease. The receptor can also be used to treat asthma,  
 CC Parkinson's disease, acute heart failure, urinary retention and  
 CC osteoporosis. In particular activators of the receptor can be used  
 CC to treat cardiovascular ailments, e.g. caused by the lack of  
 CC pulmonary blood flow or hypertension. They may also be used to  
 CC treat disorders relating to abnormal control of fluid and  
 CC electrolyte homeostasis, and in diseases associated with abnormal  
 CC angiotensin-induced aldosterone secretion. Receptor inhibitors  
 CC can be used to treat, e.g. angina pectoris, allergies, psychotic  
 CC and neurological disorders and dyskinesias. They have also been  
 CC useful in reversing endogenous anorexia, and in the control of  
 CC bulimia.  
 CC Sequence 404 AA;

Query Match 19.0%; Score 79; DB 28; Length 404;  
 Best Local Similarity 34.4%; Pred. No. 5.37e+01;  
 Matches 11; Conservative 9; Mismatches 9; Indels 3; Gaps 3;

Db 121 wgsfscrthfyfvmysiffvlclsv-dr 151  
 QY 11 WGVFKIKFSE-LY-INGYTDIVLVVFTVFER 40

RESULT 10  
 ID W34984 standard; peptide; 404 AA.  
 AC W34984;  
 DT 22-MAY-1998 (first entry)  
 DE Human G-protein coupled receptor.  
 KW Human; G-protein coupled receptor.  
 OS Homo sapiens.  
 PN WO9744359-A1.  
 PD 27-NOV-1997.  
 PF 20-MAY-1996; U07622.  
 PR 20-MAY-1996; WO-U07622.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PA (TAKE-) TAKEDA CHEM IND LTD.  
 PI Fujii R, Hinuma S, Li Y, Ruben SM, Soppet DR;  
 DR N-PSDB; T94184.  
 DR N-PSDB; T94184.

PT Polynucleotide encoding human G protein coupled receptor - specific  
 PT agonists and antagonists of which can be used to treat, e.g.  
 PT cardiovascular disease  
 PS Claim 7; Pages 35-37; 49pp; English.  
 CC The present sequence is a human G-protein coupled receptor,  
 CC which can be used to identify modulators or generate antibodies. The  
 CC modulators can be used to treat conditions requiring receptor  
 CC inhibition or activation, e.g. activators are used to treat asthma,  
 CC Parkinson's disease, acute heart failure, urinary retention and  
 CC osteoporosis, and inhibitors are used to treat hypertension, angina  
 CC pectoris, ulcers, benign prostatic hypertrophy, schizophrenia and  
 CC Huntington's and cardiovascular disease. The antibody is a  
 CC potential modulator, and can be used as an immunoassay reagent for  
 CC the identification of modulators. The receptor cDNA can be used for  
 CC recombinant production, while fragments can be used as diagnostic  
 CC probes and primers, and antisense sequences can be used to inhibit  
 CC gene expression. The cDNA can also be used for chromosome  
 CC identification.  
 CC Sequence 404 AA;

Query Match 19.0%; Score 79; DB 28; Length 404;  
 Best Local Similarity 34.4%; Pred. No. 5.37e+01;  
 Matches 11; Conservative 9; Mismatches 9; Indels 3; Gaps 3;

Db 121 wgsfscrthfyfvmysiffvlclsv-dr 151  
 QY 11 WGVFKIKFSE-LY-INGYTDIVLVVFTVFER 40

RESULT 11  
 ID W29683 standard; Protein; 1099 AA.  
 AC W29683;  
 DT 21-DEC-1998 (first entry)  
 DE Human Na-K-2Cl cotransporter NKCC2.  
 KW Na-K-2Cl cotransporter; NKCC2 gene; human; ion transport;  
 KW Bartter's syndrome; hypokalaemic alkalosis; hypercalciuria;  
 KW nephrocalcinosis; diagnosis; therapy.  
 OS Homo sapiens.  
 PN WO9829431-A1.  
 PD 09-JUL-1998.  
 PF 19-DEC-1997; U23553.  
 PR 31-DEC-1996; US-778052.  
 PA (UYA) UNIV YALE.  
 PI Lifton RP, Simon DB;  
 DR WPI; 98-388029/33.  
 DR N-PSDB; V40562.

PT Thiazide sensitive cotransporter, ATP sensitive K channel and NaK2Cl  
 PT cotransporter genes - useful for developing products for the  
 PT diagnosis and treatment of ion transport disorders, e.g. Gitelman's  
 PT Syndrome or Bartter's Syndrome  
 PS Disclosure; Fig 7; 105pp; English.  
 CC This is the amino acid sequence of human renal bumetanide-sensitive

CC Na-k-2Cl cotransporter NKCC2 of the thick ascending limb of Henle's  
 CC loop. Mutations in the NKCC2 gene (see V40562) cause Bartter's  
 CC syndrome, a disorder featuring salt wasting and hypokalaemic  
 CC alkalosis associated with marked hypercalcaemia and frequently  
 CC nephrocalcinosis. Different molecular variants inferred to alter  
 CC the structure of NKCC2 protein were identified in patients with  
 CC Bartter's syndrome by SSCP using 27 pairs of specific primers (see  
 CC V45677-730). Identification of the molecular basis of Bartter's  
 CC syndrome provides for the genetic diagnosis of this disorder. A  
 CC method is claimed for determining the presence or absence of a  
 CC mutated protein conferring altered ion transport by analysing  
 CC NKCC2 samples. Also claimed are vectors, host cells, a method of  
 CC producing NKCC2 protein, and an antibody that binds altered, but  
 CC not wild-type, NKCC2 protein. The products and methods can be used  
 CC for diagnosis of Bartter's syndrome, to identify carriers and to  
 CC identify modulators of NKCC2 function of therapeutic value.  
 SQ Sequence 1099 AA;

Query Match 19.08; Score 79; DB 36; Length 1099;  
 Best Local Similarity 35.7%; Pred. No. 5.37e+01;  
 Matches 15; Conservative 8; Mismatches 17; Indels 2; Gaps 2;

Db 966 frikfadihiqdirpnkeswkvfeemierpyrlhesckdl 1007

QY 14 FKIKSELINGTDTI-VLVVTFVFSAEAYVHLIS-RGL 53

# RESULT 12

ID W77749 standard; Protein; 197 AA.  
 AC W77749;  
 DT 30-OCT-1998 (first entry)  
 DE Staphylococcus aureus protein of unknown function.  
 KW Staphylococcus aureus protein; immune response induction; eye infection;  
 KW antibody production; T-cell immune response; gastrointestinal infection;  
 KW respiratory infection; inhibitor; bacterial infection; cardiac infection;  
 KW central nervous system; kidney infection; urinary tract infection;  
 KW antimicrobial compound identification; broad spectrum antibiotic;  
 KW therapy.

OS Staphylococcus aureus.  
 FH Key Location/Qualifiers  
 FT Misc\_difference 1..197  
 FT /note- "residues designated X are unspecified, and  
 represented as Xaa in the specification"

EP-841394-A2.  
 PD 13-MAY-1998.  
 PF 24-SEP-1997; 307485.  
 PR 24-SEP-1996; US-027032.  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 PA (SMIK ) SMITHKLINE BEECHAM PLC.  
 PI Black MT, Burnham MKR, Hodgson JE, Knowles DJC,  
 PI Lonetto MA, Nicholas RO, Pratt JM, Reichard RW, Rosenberg M,  
 PI Ward JM;  
 DR WPI: 98-252940/23.  
 DR N-PSDB: V53537.

PT New nucleic acid sequences from Staphylococcus aureus WCHU29 -  
 PT useful in vaccines and for treatment of bacterial infections of e.g.  
 PT respiratory tract and central nervous system  
 PS Claim 11; Page 369-370; 390pp; English.  
 CC This sequence represents a Staphylococcus aureus protein of unknown  
 CC function, and is encoded by a DNA sequence of the invention.  
 CC The DNA sequences were isolated from Staphylococcus aureus WCHU29  
 CC (NCIMB 40771). Host cells containing the DNA sequences are used to  
 CC produce polypeptides or fragments. The proteins are used in the treatment  
 CC of disease, for inducing an immune response by administering them, to  
 CC produce antibody and/or T-cell immune response. Antagonists of the  
 CC proteins are used for the inhibition of bacterial polypeptides.  
 CC Conditions which may be treated include bacterial infections, especially  
 CC respiratory, cardiac, gastrointestinal, central nervous, eye, kidney,  
 CC urinary tract, skin, bones and joints. The proteins can also be used to  
 CC identify antimicrobial compounds which are broad spectrum antibiotics,  
 CC especially useful in the treatment of H. pylori infection.  
 SQ Sequence 197 AA;

Query Match 18.8%; Score 78; DB 34; Length 197;  
 Best Local Similarity 16.2%; Pred. No. 6.36e+01;  
 Matches 6; Conservative 16; Mismatches 13; Indels 2; Gaps 2;

Db 89 kixavvqg-yddttmsif-vydvvdewilrlqpn1 123

QY 17 KFSEYIHGTYDVLVLTFFVFSAEAYVHISRGL 53

# RESULT 13

ID R71932 standard; Protein; 400 AA.  
 AC R71932;  
 DT 23-SEP-1995 (first entry)  
 DE I-branching enzyme.  
 KW Beta-1,6-N-acetylglucosaminyltransferase; I-branching enzyme; IGNT;  
 KW inflammation; antiinflammatory; tumor; hypersensitivity; anemia;  
 KW transgenic animal; I antigen.  
 OS Homo sapiens.

PN W09507020-A.  
 PD 16-MAR-1995.  
 PF 09-SEP-1993; U08476.  
 PR 09-SEP-1993; WO-U08476.  
 PA (LJOL-); LA JOLLA CANCER RES FOUND.  
 PI Bierhuizen MFA, Fukuda M;  
 DR WPI: J5-123182/15.  
 DR N-PSDB: Q89201.

PT Nucleic acid encoding I-branching enzyme - used to develop prods.  
 PT for treating e.g. inflammatory responses, tumours or  
 PT hypersensitivity reactions  
 PS Disclosure; Fig.6; 56pp; English.  
 CC Poly-A RNA isolated from human PA-1 teratocarcinoma cells was  
 CC reverse-transcribed and inserted into expression vector pcDNA1.  
 CC Plasmid DNA was used to transfect CHO-py-leu cells. Transfected  
 CC cells were screened using human anti-I antigen antibodies and goat  
 CC anti-human IgM. Plasmid DNA was analyzed to obtain DNA (given in  
 CC Q89201) encoding I-branching enzyme (R71932).  
 SQ Sequence 400 AA;

Query Match 18.8%; Score 78; DB 13; Length 400;  
 Best Local Similarity 25.08; Pred. No. 6.36e+01;  
 Matches 14; Conservative 14; Mismatches 4; Indels 0; Gaps 0;

Db 8 lfiisvsvliifvfsvfnfggdp 31

QY 21 LYIHGYTDIVLVLTFFVFSAE 44

# RESULT 14

ID R92474 standard; Protein; 400 AA.  
 AC R92474;  
 DT 22-APR-1996 (first entry)  
 DE Beta-1,6-N-acetylglucosaminyltransferase, I-branching enzyme.  
 KW Beta-1,6-N-acetylglucosaminyltransferase, I-branching enzyme; IGNT;  
 KW glycosyltransferase; blood group; I antigen; polylactosaminoglycan.

OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT domain 7:25  
 FT /label= Membrane-anchoring\_domain  
 FT modified\_site 37  
 FT /label= N-glycosylation\_site  
 FT modified\_site 212  
 FT /label= N-glycosylation\_site  
 FT modified\_site 255  
 FT /label= N-glycosylation\_site  
 FT modified\_site 314  
 FT /label= N-glycosylation\_site  
 FT modified\_site 388  
 FT /label= N-glycosylation\_site  
 PN US5484593-A.  
 PD 16-JAN-1996.  
 PF 09-SEP-1993; 118906.  
 PR 09-SEP-1993; US-118906.  
 PA (LJOL-) LA JOLLA CANCER RES FOUND.

```

Query Watch.      18.8%; Score 78; DB 29; Length 400;
Best Local Similarity 25.0%;
Matches      6; Conservative 14; Mismatches 4; Indels
DQ      8 lfiisvssvliifvsvfnfgdgp 31
      ||| : : : : : ||| : : :
QY      21 LYIHGTYDIWLVVTFVTFERSAEA 44

```

Search completed: Wed Dec 22 10:28:45 1999  
Job time : 18 secs.

Query Match	18.8%;	Score 78;	DB 16;	Length 400;
Best Local Similarity	25.0%;	Pred. No. 6.36e+01;		
Matches	6;	Conservative 14;	Mismatches 4;	Indels 0; Gaps 0;

```

Db      8  lfisvssvlfivsvfnfggdp 31
        | : : : : : | : : : :
QV     21  LYHGYTDIVLVVFTVFERSAEA 44

```

[illegible]

US5731420-A.  
24-MAR-1998.  
PD  
PF 07-JUN-1995; 486196.  
PF 09-SEP-1993; US-118906.  
PR 07-JUN-1995; US-486196.  
PR (LJOL-) LA JOLLA CANCER RES FOUND.  
PA Bierhuizen MFA, Fukuda M;  
PI WPI: 98-216555/19.  
DR N-PSDB; V16000.  
DR Antibodies to human I-bran  
PT beta-1,6-N-acetyl-glucosaminyl-transferase - useful for, e.g.  
PT imaging IgNt-containing sites in vivo as IgNt is involved in tumour  
PT cell adhesion to endothelium  
PS Claim 3; Fig 6A-B; 30pp: English.  
CC This is a human beta-1,6-N-acetylglucosaminyltransferase, an I-bran  
CC enzyme (IgNt). The IgNt nucleic acid encodes both the membrane-bound and  
CC soluble IgNt. An antibody or an antigen-binding fragment is specifically  
CC reactive with IgNt or an active fragment comprising a catalytic domain of  
CC IgNt. These anti-IgNt antibodies can be used to detect IgNt or IgNt-  
CC expressing cells in biological samples, e.g. to probe human CDNA  
CC expression libraries, to purify IgNt from compositions containing it, or  
CC to image IgNt-containing sites in vivo as IgNt is involved in tumour cell  
CC adhesion to endothelium and leukocyte adhesion to inflammatory sites.  
SQ -Sequence 400 AA;